Policy Issues Associated with Undertaking a Large U.S. Population Cohort Project on Genes, Environment, and Disease

A Draft Report of the
Secretary's Advisory Committee on Genetics, Health, and Society
May 2006

Chair

Reed V. Tuckson, M.D. (2007)
Senior Vice President
Consumer Health & Medical Care Advancement
UnitedHealth Group
MN 008-T902
9900 Bren Road East
Minnetonka, MN 55343

Members

Sylvia Mann Au, M.S., C.G.C. (2009)
Hawaii State Genetics Coordinator
Genetics Program
Hawaii Department of Health
741 Sunset Avenue
Honolulu, HI 96816

Cynthia E. Berry, J.D. (2007)
Partner
Powell Goldstein LLP
901 New York Avenue, NW
Third Floor
Washington, DC 20001

Chira Chen (2007)
1315 5th Avenue
San Francisco, CA 94122

James P. Evans, M.D., Ph.D. (2009)
Associate Professor of Genetics and Medicine
Director of Clinical Cancer Genetics and the
Bryson Program in Human Genetics
Departments of Medicine and Genetics
University of North Carolina at Chapel Hill
Chapel Hill, NC 27599-7624

Andrea Ferreira-Gonzalez, Ph.D. (appt pending)
Professor of Pathology and
Director of Molecular Diagnostics Laboratory
Virginia Commonwealth University
403 North 13th St
Clinical Support Center Room 246
Richmond, VA 23298-0248

Kevin T. Fitzgerald, S.J., Ph.D. (2006)
Dr. David P. Lauler Chair in Catholic Health Care Ethics
Research Associate Professor
Department of Oncology
Georgetown University Medical Center
4000 Reservoir Road, NW
Washington, DC 20057

Debra G.B. Leonard, M.D., Ph.D. (2006)
Vice-Chair of Laboratory Medicine
New York Presbyterian Hospital, Cornell Campus
Room F715, Mailbox 79
525 East 68th Street
New York, NY 10021

Julio Licinio, M.D. (appt pending)
Professor & Chairman
Miller School of Medicine University of Miami
Department of Psychiatry & Behavioral Sciences (D-28)
UM/JMH Mental Health Hospital Center
1695 NW 9th Avenue, Suite 3100
Miami, Florida 33136

Agnes Masny, R.N., M.P.H., M.S.N. (2006)
Adjunct Assistant Professor of Nursing
Temple University College of Allied Health Professionals
Research Assistant & Nurse Practitioner
Family Risk Assessment Program
Fox Chase Cancer Center
7701 Burholme Avenue
Philadelphia, PA 19111

Barbara Burns McGrath, RN, Ph.D. (appt pending)
Research Associate Professor
School of Nursing
Box 357263
University of Washington
Seattle, Washington 98195

Joseph Telfair, Dr.P.H., M.P.H., M.S.W. (2008)
Associate Professor
Department of Maternal and Child Health
School of Public Health
University of Alabama at Birmingham
1665 University Boulevard, Room 320
Birmingham, AL 35294-0022

Steven Teutsch, M.D., M.P.H. (appt pending)
Executive Director
Outcomes Research & Management
Merck & Co., Inc.
PO Box 4, WP39-130
West Point, PA 19486-0004

Huntington F. Willard, Ph.D. (2007)
Director
Institute for Genome Sciences & Policy
101 Science Drive, CIEMAS 2379
Duke University
Box 3382
Durham, NC 27708

Emily S. Winn-Deen, Ph.D. (2006)
Vice President
Strategic Planning and Business Development
Cepheid
904 Caribbean Drive
Sunnyvale, CA 94089

Ex Officios

Administration for Children and Families

Mr. Martin Dannenfelser
Deputy Assistant Secretary for Policy & External Affairs
Administration for Children & Families
370 L'Enfant Promenade, SW, Suite 600-W
Washington, DC 20447

Agency for Healthcare Research and Quality

Francis D. Chesley, Jr., M.D.
Director
Office of Extramural Research, Education, and Priority Populations
Agency for Healthcare Research & Quality
540 Gaither Road
Rockville, MD 20850

Centers for Disease Control and Prevention

Muin J. Khoury, M.D., Ph.D.
Director
Office of Genomics & Disease Prevention
Centers for Disease Control & Prevention
4770 Buford Highway, Mail Stop K-89
Altlanta, GA 30341

Centers for Medicare & Medicaid Services

Barry M. Straube, M.D.
Acting Chief Medical Officer
Acting Director
Office of Clinical Standards & Quality
Centers for Medicare & Medicaid Services
7500 Security Boulevard, MS S3-02-01
Baltimore, MD 21244-1850

Department of Commerce

Willie E. May, Ph.D.
Director
Chemical Science and Technology Laboratory
National Institute of Standards and Technology
100 Bureau Drive
MS 8300, 227/A311
Gaithersburg, MD 20899-8300

Department of Defense

Lt. Col. Scott D. McLean, MC, USA
Chief of Medical Genetics
59MDW/MNP Wilford Hall Medical Center
Department of Defense
2200 Bergquist Drive, Suite 1
Lackland AFB, TX 78236-5300

Department of Energy

Daniel Drell, Ph.D.
Biologist
Life Sciences Division
Office of Biological and Environmental Research
Department of Energy
1000 Independence Avenue, SW
Washington, DC 20585-1290

Department of Justice

TBD

Department of Labor

Thomas Alexander, J.D.
Chief of Staff
Employee Benefits Security Administration
Department of Labor
200 Constitution Avenue, NW
Washington, DC 20210

Department of Veterans Affairs

Ellen Fox, M.D.
Director
National Center for Ethics in Health Care
Department of Veterans Affairs
810 Vermont Avenue, NW
Washington, DC 20420

Equal Employment Opportunity Commission

The Honorable Cari M. Dominguez
Chair
Equal Employment Opportunity Commission
1801 L Street, NW
Washington, DC 20507

Federal Trade Commission

Matthew Daynard, J.D.
Senior Attorney
Bureau of Consumer Protection
Division of Advertising Practices
Federal Trade Commission
601 New Jersey Avenue, NW, Room 3221
Washington, DC 20580

Food and Drug Administration

Steven Gutman M.D., M.B.A.
Director
Office for In Vitro Diagnostic Device Evaluation and Safety
Food and Drug Administration
2098 Gaithersburg Road
MSC HFZ 440
Rockville, MD 20850

Health Resources and Services Administration

Denise Geolot, Ph.D., R.N.
Director
Center for Quality
Health Resources and Services Administration
5700 Fishers Lane
Rockville, MD 20857

National Institutes of Health

Francis S. Collins, M.D., Ph.D.
Director
National Human Genome Research Institute
Building 31, Room 4B09
31 Center Drive, MSC 2152
Bethesda, MD 20892

Office for Civil Rights

Robinsue Frohboese, J.D., Ph.D.
Principal Deputy Director
Office for Civil Rights
200 Independence Avenue, SW, Room 515F
Washington, DC 20201

Office for Human Research Protections

Michael A. Carome, M.D.
Associate Director for Regulatory Affairs
Office for Human Research Protections
1101 Wootton Parkway, Suite 200
Rockville, MD 20852

Office of Public Health and Science

Cristina Beato, M.D.
Principal Deputy Assistant for Health
Office of Public Health and Science
Department of Health and Human Services
200 Independence Avenue, S.W.
Hubert H. Humphrey Bldg, Room 716G
Washington, DC 20201

Table of Contents

EXECUTIVE SUMMARY AND POLICY OPTIONS……..……..…4

I. INTRODUCTION……..…….11

II. SCIENTIFIC BACKGROUND…….……...……15

- Methods for Identifying the Genetic Basis of Disease
- Biobanks and Large Population Studies
- Overview of a Hypothetical U.S. Large Population Cohort Project for the
- Study of Genes, Environment, and Health

III. THE NEED FOR PUBLIC ENGAGEMENT……...21

IV. POLICY ISSUES ASSOCIATED WITH A LARGE POPULATION COHORT PROJECT FOR THE STUDY OF GENES, ENVIRONMENT, AND COMMON DISEASE……23

- Issues Related to Research Policies
  - Need for and Value of Such a Project
    - Arguments in Favor of a Large Population Project
    - Arguments That Favor Pooling of Existing Cohort Studies and Biobanks
    - Arguments That Favor a Combination of Approaches
    - Arguments That Question the Value of a Large Population Project
  - Cost and Effects on Other Areas of Science
  - Capacity to Conduct Interdisciplinary Science
  - Need for Partnerships
  - Access to Data and Materials and Intellectual Property Concerns
- Options for Addressing Research Policy Issues
- Issues Related to Research Logistics
  - Enrollment Criteria and Recruitment of Racial/Ethnic Groups
    - Race, Ethnicity, and Sex
    - Recruiting
  - Measuring and Understanding Differences in Health and Risk Factors in the Population
  - Coordination across Multiple Institutions and Healthcare Systems
- Options for Addressing Research Logistics Issues
- Issues Related to Regulatory and Ethical Considerations
  - Institutional Review Board Review
  - Informed Consent
  - Providing Care and the Therapeutic Misconception
  - Privacy and Confidentiality
  - Control of Samples and Data

- Returning Research Results
- Options for Addressing Regulatory and Ethical Considerations
- Issues Related to Public Health Implications of the Project
- Options for Addressing the Public Health Implications of the Project
- Issues Related to Social Implications of the Project
  - Elucidating and/or Exacerbating Health Disparities
  - The Risks of Genetic Determinism
  - Developing Reasonable Social and Policy Responses to Research Findings
- Options for Addressing the Social Implications of the Project

V. OPTIONS FOR ENGAGING THE PUBLIC……..……..48

- NHGRI Public Consultation Initiative
- Options for Public Engagement

VI. CONCLUSION……...……..53

APPENDICES……..54

A. International Biobanking Efforts

- U.K. Biobank
- Biobank Japan
- Estonian Genome Project
- Icelandic Genetics and deCODE

B. Public Consultation in International Projects

- deCODE and the Icelandic Healthcare Database
- Estonian Genome Project
- U.K. Biobank
- CARTaGENE

EXECUTIVE SUMMARY AND POLICY OPTIONS

Introduction

Characterizing human genetic variation and how genetic variants interact with environmental factors (physical, behavioral, and social)[1] to influence health is currently one of the most pressing goals for scientists trying to unravel and understand the underlying causes of common diseases. Scientists hope that major clinical and public health advances will be realized by learning where variation among individuals lies within the genome, how it differs among healthy, predisposed, and sick individuals, and how particular variants of DNA interact with each other and diverse environmental factors. Large longitudinal population studies, involving the collection of data about and biological specimens from hundreds of thousands of people, offer one promising approach to learning more about the relationships among genes, the environment, and common disease. The creation of such a large database and biobank could serve as an essential research resource for hundreds, if not thousands, of research studies. For many, such a large-scale project[2] is a logical next step following the complete sequencing of the human genome. In the United States, the National Institutes of Health (NIH) is investigating the possibility of mounting a large population cohort project.

In 2004, the Secretary's Advisory Committee on Genetics, Health, and Society (SACGHS), an advisory committee to the Department of Health and Human Services (HHS), created a Task Force on Large Population Studies to gather information on the issues involved in undertaking a large population project. In addition, NIH Director Dr. Elias Zerhouni asked SACGHS to identify key policy issues related to a potential large-scale project and provide advice on what scientific**,** public, and ethical processes and approaches NIH or HHS might use in making optimal decisions about undertaking such an effort.

A large population research project raises multiple policy issues because 1) it will involve an unprecedented number of participants and, thereby, will have a significant public profile and a direct impact on many people; 2) it requires a relatively large investment of public resources and, as such, warrants scrutiny of and deliberation about its relative value to science, society, and the Nation; and 3) the nature of the information that will be derived from it raises ethical, legal, social and public policy concerns that could be unique and/or significant, particularly in view of the number of potential participants.

This report summarizes SACGHS's findings and conclusions relevant to the development of a large population research project in the United States. It focuses on preliminary and intermediate

[1] The term "environment" is used broadly as it relates to human health. For example, the World Health Organization defines environmental health as those aspects of human health, including quality of life, that are determined by physical, chemical, biological, social, and psychosocial factors in the environment. It also refers to the theory and practice of assessing, correcting, controlling, and preventing those factors in the environment that can potentially affect adversely the health of present and future generations. See www.who.int/phe/en/.

[2] Such projects have been referred to in the singular, as a study (e.g., Collins, F.S. [2004] The case for a U.S. prospective cohort study of genes and environment. *Nature*. 429:475-477); as databases (e.g., the Icelandic Health Sector Database); and as "biobanks" or research resources (e.g., U.K. Biobank). In this report, SACGHS uses the term "project" to refer to an effort that would involve the longitudinal collection and storage of data and biological specimens from large numbers of people for the research use of multiple investigators and investigative teams.

questions, steps, and strategies in five areas that need to be addressed before considering the larger question of whether the United States should undertake such a project: research policy; research logistics; regulatory and ethical considerations; the public health implications of the project; and the social implications of the project.

The Need for Public Engagement

In Committee discussions, SACGHS concluded that, in general, the processes by which large research funding decisions are made have served the public interest. Public accountability and leadership have been and continue to be key aspects of NIH's and other funding agencies' stewardship of the biomedical, public health, and behavioral research enterprise. This is essential to maintaining public trust,[3] reassuring Congress that the public's interest is being served, and ensuring that the tactical and strategic objectives for research missions are thoughtfully selected, effectively pursued, and responsive to national health concerns.

Options for Engaging the Public

With the growing enterprise of clinical and population research has come the need to inform the public about the underlying science, engage the public in discussions of priorities for federal research spending, and seek support for important areas of research. However, new issues with strong scientific content sometimes seem particularly ill-suited to one-time techniques for soliciting opinion (e.g., a typical opinion poll).[4] Because most members of the public will be unfamiliar with the concepts of a large population project, concerted efforts must be made to educate, inform, and solicit feedback and input. In the last 10 to 15 years, increasing efforts to consult lay people about scientific issues have produced a range of new methods for doing so.

Throughout this report, SACGHS has suggested several options for engaging the public in discussions and decisions about undertaking a large population project, including the importance of consulting with the scientific and international communities, representatives of populations that might be involved in the research, healthcare providers and their institutions, and those who volunteer to participate in the project as research subjects. The Committee encourages that efforts be made at all levels to develop a broader understanding of the issues involved so that they can be identified early in the process and addressed fairly and responsibly both before and throughout the duration of the proposed project.

1. **The public's willingness to participate in a large population project should be assessed before embarking on such an expensive endeavor. Willingness could be assessed through opinion polls, requests for comments posted on agency websites, or through other measures. Such an assessment should be made in advance of a funding decision.**

[3] Public Trust in Clinical Research (2005). Report and Recommendations of the NIH Director's Council of Public Representatives. Available at http://copr.nih.gov/reports/public_trust_clinical_research.pdf.
[4] *Information and Attitudes: Consulting the Public About Biomedical Science* (2005). A report published by the Wellcome Trust.

2. **If a decision is made to proceed with the project, it will be important to ensure that public engagement occurs throughout all aspects and stages of the research process, from conceptualization through design, planning, implementation, conduct, and data analysis and reporting. Public engagement also will be important in applying the knowledge gained by the research and in addressing its implications. The Secretary should ensure that sufficient project resources are dedicated to public consultation activities both before and throughout the duration of the project.**

Issues Related to Research Policy

There is a diversity of expert opinion within the scientific community about the wisdom of proceeding with a population project of this magnitude and cost at this time. For example, although a large cohort project may be needed to collect sufficient data to elucidate the contribution of genetic variation and environmental factors to common diseases, some believe it may not necessarily lead to a better understanding of common diseases or population health benefits if it does not include a carefully designed, hypothesis-driven, disease-specific component. Others believe that such a project cannot be hypothesis driven, but rather that it should be a viewed as a data and tissue resource for researchers to mine.

Some believe we have not yet made a sufficient investment in refining the methods to measure the true dynamic interaction of genes and exposures. The concerns of others focus on the ripple effects such a costly study might have on other research areas or funding opportunities.

Additionally, questions concerning fair access to data and samples have risen as the possibility of this project evolves. Another concern is the effect that studies of this caliber could have on intellectual property rights. Other questions arise about the need for collaboration with international efforts and the appropriate role of the private sector. And, given the many scientific and academic disciplines that will be required to develop such a complex and broad-based project—and the need for involvement of teams of experts in human genetics, medicine, behavior, public health, sociology, epidemiology, and environmental health science—there may be insurmountable challenges to fostering a multidisciplinary team approach. In considering the issues involved in undertaking such a large-scale research project, the Secretary should ensure that there is widespread consultation about its merit and implementation with the U.S. scientific community, HHS agency leadership, the international community, and policymakers in Congress.

Options for Addressing Research Policy Issues

3. **The HHS Secretary, in consultation with the NIH Director, should ensure that there are opportunities available to the general scientific community to a) be informed about the potential for such a project; b) present its views about the scientific validity and feasibility of such a project; c) present its views on the commitment of resources to such an effort, including whether there are benefits to leveraging existing efforts; and d) provide input on issues related to fair access by scientists to the resources and the sharing of data and samples.**

4. **Given the transdisciplinary nature of the project and its potential scope, the Secretary may wish to establish a highly collaborative model of project leadership and management in multiple HHS agencies that includes biological, behavioral, social, public health, and population-scientific disciplines as well as basic biological scientists and epidemiologists.**

5. **The HHS Secretary should continue to promote and facilitate ongoing consultation with the international community and the private sector to explore opportunities for collaboration.**

6. **In embarking on such a large-scale project, the HHS Secretary, in consultation with the NIH Director, other HHS agencies, and appropriate congressional committees, should ensure that there is widespread support for sustaining a long-term and stable investment in a large population project.**

7. **To ensure that the public benefits from such discoveries, the Secretary should require that there be clear intellectual property policies in place for discoveries made using the data and samples collected through the project.**

Issues Related to Research Logistics

Beyond the specific design issues of a large-scale project are logistical considerations that could have social and ethical consequences. These considerations include developing enrollment and data collection procedures that accurately and fairly capture the ethnic, racial, and socioeconomic diversity of populations; coordinating across multiple healthcare systems that lack universal, electronic medical record keeping systems; coordinating the multitude of healthcare institutions involved with the enrollment and data collection components of the project; creating new databases for data storage; defining and harmonizing the variables to be collected; and developing technologies to accurately and unobtrusively collect environmental information of all types, including that involving the physical, behavioral, and social environments. These considerations warrant public consideration and input.

Options for Addressing Research Logistics Issues

8. **To ensure diversity and appropriate representation in the population to be studied, the HHS Secretary should encourage project leadership and the scientific community to develop clear, consistent definitions and parameters for the stratification and classification of the projected sample population.**

9. **To ensure that subject selection is fair and just, the HHS Secretary should seek input from the public as well as researchers and clinicians on the best approaches to identifying subpopulations for recruitment, as well as issues to be considered in approaching, educating, and enrolling various subpopulations. Project organizers could be encouraged to consult with community-based organizations as one of many appropriate recruitment and enrollment strategies.**

10. **To refine methods for collecting and analyzing environmental (physical, behavioral, and social) factors influencing health, the HHS Secretary, in consultation with the NIH Director, should ensure that resources are devoted to developing new tools to validate existing methods, as well as to improving assessments of the environment, as broadly defined.**

11. **To develop uniform and secure approaches for collecting, storing, tracking, and centralizing clinical information to be gathered over the course of the project—including the use of electronic medical records—the HHS Secretary should encourage project leadership to consult with healthcare providers and organizations.**

Issues Related to Regulatory and Ethical Considerations

Many of the ethical and regulatory issues likely to be relevant to such a large-scale project are not necessarily unique or new. For example, ensuring independent ethics review of research protocols, obtaining informed consent of subjects, and protecting privacy and confidentiality—while clearly relevant to a large population project—are important issues in all clinical research studies. Nonetheless, the size and magnitude of the project could either amplify or mask ethical concerns. It will be important for collection sites, data and specimen managers, and investigators to conduct activities consistently and uniformly and in accordance with all ethical and regulatory requirements.

Options for Addressing Regulatory and Ethical Considerations

12. **To ensure that all research sites involved in the project are aware of and implement the regulations established to protect research subjects, medical privacy, and patient safety, the HHS Secretary, in consultation with the NIH Director, should convene a working group of representatives from the Office for Human Research Protections, the Food and Drug Administration, the Office for Civil Rights, relevant HHS agencies, and the Institutional Review Board and scientific communities to develop a set of recommended best practices and standard operating procedures for the project. Public input on the policies and procedures also can be sought.**

13. **To ensure that the appropriate protections of subjects' rights and welfare are in place and are being consistently implemented, project leadership should systematically and regularly seek the input of study subjects regarding their experiences, concerns, and recommendations for enhancing protections.**

14. **To promote the ethical use of clinical and epidemiological data and specimens collected through the project, project leadership could develop guidance on how such data and samples can be used and under what conditions. This guidance should be made available to project participants so that they are informed of the protections that are in place and that are to be expected.**

Issues Related to Public Health Implications of the Project

Some argue that the genetic science of common complex diseases is simply not mature enough to provide a profound understanding of the reasons why the genetic factors being identified as significant in one study cannot be replicated in another. Moreover, it is difficult in many cases to move from a statistical genetic association to an understanding of the mechanism of action that would suggest new therapies or preventive measures or that would withstand evidence-based regulatory decisionmaking. This leads to questions about whether large population cohort studies actually can provide results that are sufficiently definitive to lead to clinical applications and whether the data gathered can be reliably extrapolated across the entire population. Without the ability to identify gene function, there also is the risk that genes or single nucleotide polymorphisms will be associated with disease, but we will not know with what certainty. Or, the association will be clear but there is no available treatment. This gap between identifying risk and providing treatment is troublesome, particularly because of its uncertain duration.

Option for Addressing the Public Health Implications of the Project

15. To advance the application of research findings resulting from the project to improve health, the HHS Secretary and project leadership should systematically and regularly disseminate study findings as they emerge from the project, with clear descriptions of the possible clinical implications of the results and the limitations of the data, their generalizability, and their clinical and public health implications. This information should be tailored to meet the information needs of the public, healthcare providers, and the public health community.

Issues Related to the Social Implications of the Project

A large population project has the potential to either exacerbate or help explain and eliminate health disparities. In addition, project leadership must exercise caution in interpreting the results of studies emanating from the project, lest they reinforce overly simplistic or deterministic explanations of disease. Social policies developed in response to research findings, for example, with regard to environmental policy, health insurance decisions, and risk assessment, must be undertaken with sufficient knowledge and deliberation.

Option for Addressing the Social Implications of the Project

16. To periodically assess persistent and emerging social implications of the project and research results, the HHS Secretary, in consultation with project leadership, should establish an independent standing committee for the duration of the project. The committee could consist of individuals with expertise in the relevant sciences, medicine, law, ethics, and patient and community advocacy. The committee should routinely seek public input on the implications of the research resulting from the project and report its findings.

Conclusion

SACGHS's goal is to help illuminate a pathway for the HHS Secretary's assessment of the merit, utility, and feasibility of a large population project. Despite the considerable challenges identified in this report, SACGHS is enthusiastic about the concept of mounting a large population project for the study of genes, environments, their interactions, and common diseases in the United States because of its potential to generate significant health benefits. However, the Committee encourages efforts to be made at all levels to develop a broader understanding of the issues involved so that they can be identified early in the process and addressed fairly and responsibly throughout the duration of the proposed project.

I. INTRODUCTION

The Secretary's Advisory Committee on Genetics, Health, and Society (SACGHS) was established in 2002 by the Secretary of Health and Human Services (HHS) as a public forum for deliberating on the broad range of human health and societal issues raised by advances in genetics and, as warranted, providing advice on these issues. In a March 2004 priority-setting process, SACGHS determined that the opportunities and challenges associated with conducting large population studies aimed at understanding the relationships of genes, environments,[5] their interactions, and common, complex diseases warranted in-depth study.

Characterizing human genetic variation and how genetic variants interact with physical, physiological, behavioral, and social environmental factors to produce disease currently is one of the most pressing goals for scientists who are trying to unravel the underlying causes of common disease. Scientists hope that major public health advances will be realized by learning where variation among individuals lies within the genome, how it differs among healthy, predisposed, and sick individuals, and how particular variants of DNA interact with each other and with environmental factors. Large longitudinal population cohort projects[6] involving the collection of data about and biological specimens from hundreds of thousands of people offer one promising approach to learning more about the relationship among genes, environments, their interactions, and common disease. The creation of such a large database and biobank could serve as an essential research resource for hundreds, if not thousands, of research studies. For many, such a large-scale effort is a logical next step following the complete sequencing of the human genome.

Currently, a number of countries have begun such national population research projects. These efforts capitalize on genome-wide scanning for single nucleotide polymorphisms (SNPs) and haplotypes that could provide population-based information about associations between common polymorphisms and common diseases. However, international experiences in designing and implementing these efforts have demonstrated the need to proceed with careful deliberation and public input.

In the United States, the National Institutes of Health (NIH) is investigating the possibility of mounting a large population cohort project. U.S. investigators already have conducted or are conducting many smaller-scale studies to detect some associations between environmental factors, genetic and biobehavioral markers, and disease. Although these studies are important and informative on their own, it is not clear whether they have the statistical power needed to definitively identify associations.

---

[5] The term "environment" is used broadly as it relates to human health. For example, the World Health Organization defines environmental health as those aspects of human health, including quality of life, that are determined by physical, chemical, biological, social, and psychosocial factors in the environment. It also refers to the theory and practice of assessing, correcting, controlling, and preventing those factors in the environment that can potentially affect adversely the health of present and future generations. See www.who.int/phe/en/.

[6] Such projects have been referred to in the singular, as a study (e.g., Collins, F.S. [2004] The case for a U.S. prospective cohort study of genes and environment. *Nature*. 429:475-477); as databases (e.g., the Icelandic Health Sector Database); and as "biobanks" or research resources (e.g., U.K. Biobank). In this report, SACGHS uses the term "project" to refer to an effort that would involve the longitudinal collection and storage of data and biological specimens from large numbers of people for the research use of multiple investigators and investigative teams.

As a first step toward honing in on associations between genes and the environment in the initiation and progression of disease, in February 2006 HHS Secretary Michael O. Leavitt announced two large-scale efforts: 1) the Genes and Environment Initiative (GEI), a research effort at NIH to combine a type of genetic analysis and environmental technology development in order to understand the causes of common diseases and 2) the Genetic Association Information Network, a public-private partnership between NIH, the Foundation for the National Institutes of Health, and Pfizer and Affymetrix, to conduct laboratory studies to determine the genetic contributions to five common diseases. These projects will rely on existing cohorts using a case-control method; that is, they will study people who have a disease in comparison to those who do not have the disease.

In addition to these new initiatives, planning has been under way for other possible large-scale efforts. The National Children's Study (NCS) (see Box A) is designed to focus on deciphering the influence of environmental exposures on childhood disease and development. The Department of Veterans Affairs (VA) also has been considering a large-scale project in clinical genomic medicine. On March 16, 2006, VA announced the formation of a committee to advise the department on emerging issues in genomic medicine. The new Genomic Medicine Program Advisory Committee will help VA establish policies for using genetic information to optimize the medical care of veterans and to enhance the development of tests and treatments for diseases that are particularly relevant to veterans. The committee is expected to meet up to three times annually and is being asked to recommend policies to gather and use both genetic and other medical information for medical care and research. In this regard, it will help lay the groundwork for the future development of a comprehensive genomic medicine program for VA.

While discussions are under way in the U.S. scientific community about the need for and possible design of such studies, SACGHS believes it is an opportune time to identify the associated social, legal, ethical, and policy issues that would be involved and to define the processes that are needed to adequately address both the scientific and societal issues. Questions that could be asked about a particular study in these areas include the following:

- How will such a project affect other areas of science and the broader scientific community?
- Which scientific disciplines must be included from the project's outset in order to ensure the best validation or development of measures and methods derived from the biological, sociobehavioral, public health, and population/epidemiological sciences?
- What are the predicted scientific benefits or gains of such a project?
- What is the best way to consult with and educate the public about the nature of such a project?
- How will subjects be recruited to participate, and what are the potential benefits or risks of participation?
- What are the potential social and health implications of future findings resulting from the project?
- What are the implications of project findings, and how would policymakers, researchers, clinicians, public health agencies, private industry (including insurance companies), and the general public act on the information?
- What are the implications of various project designs for outcomes?

Box A: Examples of Existing or Proposed U.S. Prospective Cohort Projects

The U.S. Framingham Heart Study

In 1948, NIH launched a prospective-cohort study of cardiovascular disease, which is the leading cause of death and illness in the United States. The Framingham Heart Study prospectively examined the heart health of more than 5,000 adults in Framingham, Massachusetts, and eventually extended the study to the children and grandchildren of the original participants. The study followed participants by providing extensive medical examinations, blood tests, and other measures of health status. The results of this longitudinal study revealed for the first time that there are risk factors that contribute to cardiovascular disease, such as cholesterol levels, high blood pressure, and diabetes. These findings dramatically altered the treatment of patients with cardiovascular disease and public education on risk factors for cardiovascular disease. The Framingham Heart Study is continuing under the auspices of NIH's National Heart, Lung, and Blood Institute (NHLBI). By studying three generations, the Framingham study is able to examine the extent to which genetic factors relate to cardiovascular disease and its risk factors.

The U.S. Multiethnic Cohort Study

In the ongoing Multiethnic Cohort (MEC) study, a large cohort of 215,000 participants from 5 distinct ethnic groups was enrolled to prospectively observe the influence of environmental and lifestyle exposures and biomarkers that are thought to alter cancer risk.[7] In this study, groups of Japanese, African Americans, Latinos, Native Hawaiians, and Caucasians are being studied to maximize the range of environmental and lifestyle exposures. The ethnic groups chosen for study differ in their relative rates of cancer and in their dietary habits. The designers of the MEC study reasoned that if members of an ethnicity have different cancer rates in different environments, it may be that environmental factors, such as diet, are contributing risk factors to disease. In addition to examining environmental factors, the MEC study is analyzing biomarkers that are believed *a priori* to alter cancer risk in order to discern their impact on disease manifestation. The biomarkers chosen were based on biological hypotheses, expression profiling, and linkage studies.

The National Children's Study (NCS)

The NCS, which has been in the planning stages for more than six years, is to be a coordinated effort between NIH (the National Institute of Child Health and Human Development and the National Institute of Environmental Health Sciences), the Centers for Disease Control and Prevention, and the Environmental Protection Agency (EPA). It will examine the lives of approximately 100,000 American children and their parents and catalog multiple environmental, social, and genetic factors related to health and disease. The study is designed to collect a wide range of environmental data to be analyzed in relationship to several specific health outcomes, including those related to pregnancy, child growth and development, injury, asthma, and psychological and emotional health. Data collection would have begun prior to participants' births and continued until their 21st birthdays. During the study, subjects will take part in a minimum of 15 in-person visits with a local research team, parents/guardians will complete additional questionnaires every 3 months until the age of 5 and annually thereafter, and biological samples will be collected for genetic analysis. Samples of the air, water, soil, and dust from the children's environments will be collected regularly. To capture America's ethnic, social, and geographic diversity, the study will enroll families from 96 locations around the country. The project is estimated to cost $2.7 billion over two decades. The future prospects of the study are uncertain; funding for the project was not included in the President's Fiscal Year 2007 budget.

[7] Kolonel, L.N., Altshuler, D., Henderson, B.E. (2004). The multiethnic cohort study: exploring genes, lifestyle and cancer risk. *Nature Reviews Cancer*. 4:1-9.

- How can we conduct such a project in a way that is fair and equitable to different subpopulations?
- What are the roles of the local, state, and federal governments and the private sector in such a project?
- Are there obstacles that would make the undertaking of such a project especially difficult in the United States compared to other countries?
- What unique ethical, legal, and regulatory factors would have to be considered, if any?

In 2004, SACGHS created the Task Force on Large Population Studies to gather information on the issues involved in undertaking a large population project. The Task Force organized a session in March 2005 to inform the full Committee about different approaches to a large population project and to facilitate a discussion of the attending scientific, logistical, ethical, legal, and social issues. Following further discussions in June 2005, SACGHS charged the Task Force with gathering additional input on these issues from members of the scientific and ethics communities at the Committee's October 2005 meeting. Dr. Elias Zerhouni, NIH Director, requested that, in addition to identifying key policy issues related to a potential large-scale project, the Committee also should provide advice on what scientific**,** public, and ethical processes and pathways might be helpful to NIH or HHS in making optimal decisions about undertaking such an effort. Dr. Zerhouni specified that the Committee could be most helpful to the Secretary by conducting an inquiry that includes the following steps:

- Step 1: Delineate the questions that need to be addressed in order for policymakers to determine whether the U.S. government should undertake, in any form, a large population project to elucidate the influence of genetic variation and environmental factors on common, complex disease.
- Step 2: Explore the ways in which, or processes by which, the questions that are identified in step 1 can be addressed, including any intermediate research studies, pilot projects, or policy analysis efforts needed.
- Step 3: Taking into account the possible ways in which the questions could be addressed, determine which approaches are optimal from a substantive and feasibility standpoint and recommend a specific course of action for moving forward.

Clearly, a large population initiative raises many policy issues because 1) it will involve an unprecedented number of participants and, thereby, will have a significant public profile and a direct impact on many people; 2) it requires a relatively large investment of public resources and, as such, warrants scrutiny of and deliberation about its relative value to science, society, and the country; and 3) the nature of the information that will be derived from it raises ethical, legal, social and public policy concerns that could be unique and/or significant, particularly in view of the number of potential participants.

This report summarizes SACGHS's findings and conclusions relevant to the development of a large population research initiative in the United States. It focuses on preliminary and intermediate questions, steps, and strategies that need to be addressed before considering the larger question of whether the United States should undertake such a project. SACGHS's goal is to help illuminate a pathway for the HHS Secretary's assessment of the merit, utility, and feasibility of a large population project. Despite the considerable challenges identified in this

report, Committee members are enthusiastic about the concept of mounting a large population project for the study of genes, environments, their interactions, and common diseases in the United States because of its potential to generate significant health benefits. The Committee welcomes the opportunity to develop a broader understanding of the issues involved.

II. SCIENTIFIC BACKGROUND

The scientific possibilities resulting from the recently completed Human Genome Project and the HapMap Project[8] (see Box B) have prompted a search for methods that will advance our understanding of the relationship between genetics and disease. Now that the sequencing of the human genome is essentially complete, scientists are left with the following major tasks: identify the genes and other functional parts of DNA that are correlated with health or disease, elucidate their functions, and gain an understanding of how they interact with the each other and with the environment. Experts in a variety of scientific fields are debating the utility and feasibility of large population studies that have the goal of elucidating the interaction of environmental exposures and genetic variation and their relationship to disease. Due to the breadth of such an initiative, an interdisciplinary approach that includes geneticists, epidemiologists, toxicologists, social and behavioral scientists, public health experts, biostatisticians, information technologists, health providers, ethicists, community representatives, and others is needed.

Methods for Identifying the Genetic Basis of Disease

Research using the sequence of the human genome has shown that any two individuals differ in their genetic makeup by only about 0.1 percent.[9] Characterizing this small fraction of variation currently is one of the most pressing goals for scientists who are trying to unravel the influence of genes on human health and disease. Personalized medicine and major advances in public health advances are expected to result from understanding the variation in DNA that makes humans different from one another in their susceptibility to disease, their physiological, mental and emotional response to physical, behavioral, and social environmental exposures, and their response to medicines. Understanding gene-environment interactions is particularly important because the recent epidemics of chronic diseases have developed over the span of a few generations. Although this is far too short a period for the genome to change dramatically, it is definitely sufficient time for substantial environmental changes to occur and to have adverse effects on those genetically predisposed to respond poorly to environmental challenges. Advances in public health and personalized medicine can be made by knowing where in the genome the variation lies, how this variation differs between healthy and sick people, and how individuals with their particular variants of DNA interact with environmental factors.

Classic genetic research methods involving family linkage analysis have been used in many instances to identify the genetic basis of simple, Mendelian (i.e., heritable) disorders. Linkage analysis examines the patterns of co-transmission of genetic markers and diseases within families through a comparison of affected and non-affected family members in order to identify

[8] The International HapMap Consortium (2005). A haplotype map of the human genome. *Nature.* 437:1299-1320.

[9] The International SNP Working Group (2001). A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature.* 409:928-933.

Box B: International HapMap Project

The haplotype map, or "HapMap," is a tool that will allow researchers to find genes and genetic variations that affect health and disease. The elucidation of the entire human genome has made possible our current effort to develop a haplotype map of the human genome.

The DNA sequence of any two people is 99.9 percent identical. The variations, however, may greatly affect an individual's disease risk. Sites in the DNA sequence where individuals differ at a single DNA base are called single nucleotide polymorphisms, or SNPs. Sets of nearby SNPs on the same chromosome are inherited in blocks. The pattern of SNPs within a block is called a haplotype. Blocks may contain a large number of SNPs, but only a few SNPs may be sufficient to uniquely identify all possible haplotype patterns within a block. The HapMap is a map of these haplotype blocks, and the specific SNPs that identify the block haplotypes are called "tag" SNPs.

The HapMap will reduce the number of SNPs required to examine the entire genome for association with a phenotype from the 10 million SNPs that exist to roughly 500,000 tag SNPs. This will make genome scan approaches for identifying the regions that contain genes that affect diseases much more efficient and comprehensive, because there will be no need to waste effort by typing more SNPs than necessary, and all regions of the genome can be included.

In addition to its utility for studying genetic associations with disease, the HapMap should be a powerful resource for studying the genetic factors that contribute to variation in response to environmental factors, in susceptibility to infection, and in the effectiveness of and adverse responses to drugs and vaccines. All such studies will be based on the expectation that there will be higher frequencies of the contributing genetic components in a group of people with a disease or particular response to a drug, vaccine, pathogen, or environmental factor than in a group of similar people without the disease or response. Using just the tag SNPs, researchers should be able to find chromosome regions that have different haplotype distributions in the two groups—those with a disease or response and those without. Each region would then be studied in more detail to discover which variants in which genes in the region contribute to the disease or response, leading to more effective preventive, diagnostic, or therapeutic interventions.

chromosomal regions that may contain disease-related genes. Efforts can then be made to zero in on candidate genes and understand their role in disease.

Classic genetic research methods involving family linkage analysis have been used in many instances to identify the genetic basis of simple, Mendelian (i.e., heritable) disorders. Linkage analysis examines the patterns of co-transmission of genetic markers and diseases within families through a comparison of affected and non-affected family members in order to identify chromosomal regions that may contain disease-related genes. Efforts can then be made to zero in on candidate genes and understand their role in disease.

Many of the common diseases that affect the American population, however, are complex and multifactorial—that is, they are caused by a complex interplay among multiple genes and environmental factors—factors in the physical environment as well as the behavioral and social environments.[10] In other words, although the presence of one or more genetic variants

[10] Palmert, M.R., Hirschhorn, J.N. (2003). Genetic approaches to stature, pubertal timing, and other complex traits. *Molecular Genetics and Metabolism.* 80:1-10.

contributes to the underlying cause of disease, it is the body's exposure to environmental factors, including behavioral and social influences, that may determine whether and how genetic variants contribute to disease manifestation. These markers of variation in biobehavioral reactivity might produce cleaner intermediate phenotypic markers of early disease vulnerability.[11]

Genetic association studies are the basis of large population studies. Such studies assess correlations between previously identified genetic variants and trait differences (such as disease status) on a population scale, rather than on a family basis.[12] These population-based studies are possible because of recent advances in identifying genetic variants in the human population, including the mapping of SNPs in individual chromosomes and even the entire genome,[13,14] the availability of high-throughput genotyping techniques, and the ability to simultaneously compare groups of genetic loci. Association studies must include large enough study populations to capture genetic variants that do not exhibit complete penetrance, but that do exhibit a significant association to particular diseases. The study population also must be large enough to detect simultaneous multiple variables that interact and cause disease, such as gene-environment or gene-gene interactions.

Population-based genetic association studies rely on samples obtained from affected and unaffected individuals. For these studies, the frequency with which certain alleles are present in each of these groups is tested for association with a disease. A common approach is to use biological information about the molecular pathology of the disease to guide the selection of candidate genes for such testing. Several sampling strategies can be used in association studies, including case-control studies and prospective cohorts. Typically, the case-control method has been used, in which genetic and environmental data are collected from persons with specific diseases or conditions and compared to those free of disease. Although case-control studies are of great value in suggesting potential etiologic factors, they cannot provide information on predictive biobehavioral markers, are prone to important biases related to case ascertainment, and often involve incomplete or biased assessment of risk modifiers or gene-environment interactions.

In comparison, prospective cohort studies of genes and environment enroll individuals prior to disease onset and prospectively collect environmental and biobehavioral marker data, allowing for the examination of contributing non-genetic and genetic factors in disease. Prospective-cohort studies must enroll more individuals than case-control studies in order to ensure that a sufficient number of affected persons within the study population eventually develop the disease of interest for evaluation and statistical analysis. By increasing the sample size, scientists increase the study's power to detect subtle differences between individuals. Prospective-cohort, large population studies are designed to find significant associations among genetic variants, traits, and environmental exposures. Collecting phenotypic and environmental information in a standardized and unbiased manner is crucial to such efforts. But even more challenging is to

[11] Moffitt, et al. (2005). Op. cit.
[12] Cardon, L.R., Bell, J.I. (2001). Association study designs for complex diseases. *Nature Reviews Genetics.* 2:91-99.
[13] Mullikin, J.C., et al. (2000). An SNP map of human chromosome 22. *Nature.* 407:516-520.
[14] Altshuler, D., Pollara, V.J., Cowles, C.R., Van Etten, W.J., Baldwin, J., Linton, L., Lander, E.S. (2000). An SNP map of the human genome generated by reduced representation shotgun sequencing. *Nature.* 407:513-516.

collect indices of variation in biobehavioral reactivity that might produce cleaner intermediate phenotypic markers of early disease vulnerability.

Large-scale cohort projects are under discussion or already under way in the United Kingdom (U.K.), Iceland, Estonia, Germany, Canada, Taiwan, and Japan (see Appendix A). Although these projects are likely to be powerful engines for research, they alone will not meet U.S. research and clinical needs for a number of reasons, including the inadequate representation of some U.S. minority subpopulation groups that bear disproportionate burdens of disease, the limited potential for research access to data and biological materials, and substantial international differences in environment, lifestyles, and healthcare. In addition, as noted earlier, several U.S. cohort projects are already under way, which enroll large, prospective cohorts for the purpose of following participants over time to evaluate the progression toward specific outcomes (often disease) (see Box A). However, the design and goals of these efforts differ from those being considered for a large-scale population project for the study of genes, environment, and health.

Biobanks and Large Population Studies

Large-scale cohort studies of different diseases would require the enrollment of a large number of individuals willing to provide research access to their specimens and medical information and would need to be able to collect information about their environmental exposures (including physical, social, and behavioral information). Data could then be stored in databases and specimens in repositories, or biobanks, which would be accessed by qualified investigators for specific research purposes (e.g., studies of specific diseases or genes of interest). More specifically, a biobank, also known as a biorepository or a genebank, is "a stored collection of genetic samples in the form of blood or tissue, that can be linked with medical and genealogical or lifestyle information from a specific population, gathered using a process of generalized consent."[15] In recent years, many biobanks have been initiated in parallel with large population studies to facilitate the simultaneous analysis of genetic material, disease status, and the environmental exposures of individuals. In some cases, the biobank is directly linked to a study with a predetermined goal, such as identifying the genes causing a specific disease. In others, the biobank literally serves as a repository of genetic material, patient exposure data, and medical history information that is available as a resource to researchers who request samples for the study of a particular disease. Characteristics of biobanks, such as participant population, age, size, ethnicity, and environmental exposures, vary widely.

The type of analyses to be used and the hypotheses to be addressed determine what kind of biological sample(s) will be collected in a biobank. At present, the types of analyses commonly used include the genotyping of markers; transcript profiling, or the measurement of how a gene or a set of genes is expressed in tissue samples; gene quantification, or the analysis of how altered copy numbers of genes and chromosomes differ between normal and malignant tissues;

[15] Austin, M.A., Harding, S. McElroy, C. (2003). Genebanks: a comparison of eight proposed international genetic databases. *Community Genetics*. 6:37-45.

and proteomic analysis, which is the analysis of protein expression and modification in response to genetic and environmental factors.[16]

For example, genotyping requires a supply of DNA and can be obtained from blood samples or from cells from the lining of the mouth cavity, provided that the samples contain cells with intact nuclei. The study of changes in genes that occur during life, such as in tumor tissues, can utilize transcript profiling and gene quantification, both of which require samples of nucleated cells from the relevant tumor and tissues. Proteomic analysis, however, does not require nucleated cell samples; rather, any body fluid or specimen related to the disease process can be collected. In many cases, the biobank samples are stored in a manner that preserves the DNA, gene transcripts such as RNA, or proteins for study at a later date. However, some biobanks have used methods that preserve the living samples in a cell culture system, which provides a renewable and permanent source of the cells and materials for analysis.

Several "national" biobanks have been created, such as the U.K. Biobank, Biobank Japan, the Estonian Genome Project (EGP), and deCODE in Iceland (as described in Appendix A). These banks differ in their design and approach to large population studies. Population diversity is a major influence on the design of large population studies. In Iceland, where the population is very homogenous and extensive genealogies are available, a larger-scale version of linkage analysis is possible. In large population studies of more diverse populations, such as those found in the U.K. and Estonia, association studies will be used to examine the population distribution of genetic variants and their association with disease.

These large population projects also differ in their ultimate objectives. In the case of the U.K. Biobank, the goal is an epidemiological analysis of risk factors that contribute to disease, while the project goal of the EGP is to maintain genetic information in a database as a resource for public health and biomedical research, as well as for the clinical management of participants. Biobank Japan aims to develop tools for personalized medicine, choosing medical procedures and drugs based on patients' genetic profiles. Ultimately, the designers of biobanks anticipate that the biobanks and their associated large population studies will become part of the research infrastructure from which future discoveries in medicine and public health can be derived. Because they are considered to be research tools and research infrastructure, they have primarily been funded by governments and non-profit organizations, although in some cases the exclusive rights to the development of therapies and diagnostics have been assigned to private companies that fund research, such as deCODE in Iceland.

Overview of a Hypothetical Large U.S. Population Cohort Project for the Study of Genes, Environment, and Health

In 2004, the Director of the National Human Genome Research Institute (NHGRI) published an essay in *Nature* asserting that rigorous and unbiased conclusions about the causes of diseases and their population-wide impact will require the conduct of a large population study over many

[16] Jonsson, L., Landegren, U. (2001). Storing and using biobanks for research. In: *The Use of Human Biobanks. Ethical, Social, Economical and Legal Aspects — Report I.* Hansson, M.G. (ed.). Uppsala University. See www.bioethics.uu.se/biobanks-report.html.

years and that the time was right for the United States to consider such a project.[17] In 2004, NHGRI, in collaboration with several other NIH institutes, commissioned a group of experts in genetics, epidemiology, biostatistics, and ethical, legal, and social issues in genetic research to examine the scientific foundations and broad logistical outlines of a hypothetical U.S. cohort project for the study of genes, environment, and health. The recommendations of this panel are summarized in the document, *Design Considerations for a Potential United States Population-Based Cohort to Determine the Relationships among Genes, Environment, and Health: Recommendations of an Expert Panel*, which was posted on the NHGRI website in June 2005.[18]

According to this group, the goal of a U.S. large population project would be ascertaining and quantifying all of the major environmental and genetic causes of common illnesses, setting the stage for a future of better preventive medicine and more effective therapy. Such an effort could have the following major characteristics:

- Representative samples of the U.S. population would be followed prospectively for the development of specific endpoints.
- The project would involve between 500,000 and 1,000,000 people.
- The project population would be sampled from defined Census tracts.
    - Population attrition is estimated optimistically at 3 percent per year and would need to be compensated through ongoing recruitment.
    - Subgroups that have not traditionally participated in research would need to be oversampled in order to ensure that the groups have sufficient numbers of individuals.
- The first year of the project would consist of public consultation, protocol development, Office of Management and Budget approval, and training.
- The project population would be recruited primarily door-to-door over a four-year period.
- Study participants (research subjects) would be contacted twice per year and the cohort re-examined on average every four years.
- Disease outcomes would be assessed using hospital records, outpatient records, and other data sources, such as Centers for Medicare and Medicaid Services (CMS) data and registries.
- Data collection at entry would include the widest breadth of phenotypes and environmental factors needed to predict outcomes, balanced by cost and participant burden.
- A core group of baseline variables would be collected in all or nearly all participants, with additional variables added to the core list for different age groups.
- Informatics and data management needs would include 1) data capture, entry, and editing; 2) database design and management; and 3) analysis.

Although funding for such an endeavor has not been identified, carefully outlining and considering the goals and key design aspects of such an initiative was deemed by this group to be

[17] Collins, F.S. (2004). The case for a U.S. prospective cohort study of genes and environment. *Nature*. 429:475-477.
[18] Available at www.genome.gov/13014436.

of high scientific importance. In May 2004, NIH sought input from the scientific community on approaches to developing a large-scale U.S. project for the study of genetic and environmental influences on common diseases. Advice could include recommendations on the optimal characteristics of such a project; recommendations on combining existing cohorts for such efforts; and characteristics of existing studies that might lend themselves to inclusion in such efforts.[19]

Furthermore, methods of public engagement will need to be considered and applied at several points along the project timeline, in order to support the very concept of such a project and to sustain public trust and interest in its continuance.

III. THE NEED FOR PUBLIC ENGAGEMENT

- Why is it necessary to consult the public about a large population project?
- At what level should public engagement about a large population project occur?
  - fundamental, conceptual level – to do or not to do?
  - project design and planning level
  - project initiation level
  - conceptual level and throughout all phases of the project
- When and how should public engagement occur?
  - as soon as possible to inform decisionmakers
  - when a funding decision about the project is made
  - after project design and planning have been completed
  - as investigators begin to access collected data and biological materials
  - when investigators have findings of clinical importance
  - throughout all phases of project proposal, planning, and conduct
- What questions should the public be asked?
- Which subgroups of the public should be engaged?
  - community level
  - racial/ethnic groups
  - broader scientific community
  - healthcare and medical communities
  - industry
  - others

Given the scope, magnitude, cost, and time span of developing a large population cohort project, the significance of research findings resulting from analyses of the data/specimens, and the need for broad public support and participation, the public must be consulted about the project's value, design, implementation, and application of research results. SACGHS believes that decisionmaking in a democratic society should take account of public attitudes, and, therefore, public engagement must be central in planning for and implementing a large population project. Some of the important questions that need to be considered about the public consultation include

[19] A summary of the advice received is available at www.genome.gov/13014436.

those involving what questions the public should be asked to address and when; what the scope of the public engagement efforts should be; and what challenges may exist to public engagement.

In Committee discussions, SACGHS concluded that, in general, the processes by which large research funding decisions are made have served the public interest. Public accountability and leadership have been and continue to be key aspects of NIH's and other funding agencies' stewardship of the biomedical and behavioral research enterprise. This is essential to maintaining public trust, reassuring Congress that the public's interest is being served, and ensuring that the tactical and strategic objectives for research missions are thoughtfully selected, effectively pursued, and responsive to national health concerns. The process for making such a decision involves many steps at which public input can play an important role, for example, through representative democracy and the advisory committee system used by several funding agencies to assist in setting priorities and making funding decisions.

Prior to assessing enthusiasm for such a project, the public's general understanding of genetics and genetic research will need to be determined. Survey data presented to SACGHS showed that respondents' awareness of genomics is very broad—with 75 percent having had some exposure to genomics—but is not very detailed, and is generally restricted to a basic understanding that there is a relationship between genes and health. In addition, attitudes toward genomics and personalized medicine are very favorable, and interest in using genetic information to understand and optimize health and make informed choices about prescription drugs is high. Slightly more than half of the American public has a favorable attitude toward using genetic information to personalize and optimize health.[20]

The survey data also suggest that even those who hold a favorable attitude toward using genetic information for health purposes have significant concerns about the privacy of their genetic information, including its storage for research purposes, and about the potential for its misuse. The public also may have misgivings about a government-sponsored DNA databank. In one recent survey, only one-quarter to one-third of the public agreed that the government should create a national database of DNA information to advance health.[21] Moreover, it will be critical that the public's willingness to participate in such a project be assessed before embarking on such an expensive endeavor. Willingness could be assessed through opinion polls, requests for comments posted on agency websites, or through other measures. Such an assessment should be made in advance of a funding decision, but might best be preceded by or combined with a public education effort about the goals, purpose, benefits, and costs of a long-term cohort project.

Consistent with its steadfast belief that public engagement will be critical before proceeding with a large population study, as well as during its implementation if a decision is made to proceed, throughout this report SACGHS will make recommendations about other possible ways to engage the public in such discussions and about the questions that might be asked and when they should be asked.

[20] A Cogent Research web-based survey of a random sample of 1,000 Americans over the age of 18. Presented to SACGHS March 28, 2006.
[21] Ibid.

The Committee also makes two overarching recommendations about the need for efforts to be made at all levels to develop a broader understanding of the issues involved so they can be identified early in the process and addressed fairly and responsibly, both before and throughout the duration of the proposed project.

1. **The public's willingness to participate in a large population project should be assessed before embarking on such an expensive endeavor. Willingness could be assessed through opinion polls, requests for comments posted on agency websites, or through other measures. Such an assessment should be made in advance of a funding decision.**

2. **If a decision is made to proceed with the project, it will be important to ensure that public engagement occurs throughout all aspects and stages of the research process, from conceptualization through design, planning, implementation, conduct, and data analysis and reporting. Public engagement also will be important in applying the knowledge gained by the research and in addressing its implications. The Secretary should ensure that sufficient project resources are dedicated to public consultation activities both before and throughout the duration of the project.**

The basis for these recommendations is discussed in more detail in Section V. At its October 2005 meeting, SACGHS heard input from several panelists about the need to consider several possible mechanisms for soliciting public input on the design, conduct, and possible clinical, public health, and social implications of such a project. These approaches are described in greater detail in Section V. In addition, an initiative of NHGRI to fund a pilot project to gather wide public input to inform the design and implementation of one or more possible large U.S. population-based studies, including a longitudinal cohort study, of the role of genes and environment in health and disease is discussed in more detail in Section V.

IV. POLICY ISSUES ASSOCIATED WITH A LARGE POPULATION COHORT PROJECT FOR THE STUDY OF GENES, ENVIRONMENTS, AND COMMON DISEASE

Most experts agree that association studies will help shed light on the genetic and environmental (physical, behavioral, and social) factors that contribute to complex phenotypes and diseases in humans, although there are many different views regarding the ultimate value of such studies and the scientific and logistical aspects of study design, particularly cohort composition, cohort size, the collection of environmental data, and the statistical analysis of resulting studies. Logistical considerations also include funding sources, the effects of funding on support for other areas of science, access to data, enrollment procedures, the need for informed consent and standardized electronic medical records, and benefit sharing. In addition, ethicists have raised concerns about the potential drawbacks inherent in studies that focus on human genetic variation, specifically those in the realm of genetics and race. In addition, there are questions involving whether such knowledge will exacerbate health disparities or promote genetic determinism or discrimination and whether the benefits will be justly shared across society.

Based on a review of current similar efforts that are under way in the United States and abroad and on the input provided by NIH on plans and issues for a possible future effort, SACGHS identified five broad policy areas requiring further consideration:

- research policies;
- research logistics;
- regulatory and ethical considerations;
- public health implications of the project; and
- social implications of the project.

Within each broad policy area, there are a number of specific policy issues and questions that will need to be addressed. Some of these issues are of more immediate concern than others. In keeping with the Committee's charge, approaches that could be employed to address these issues are suggested. Other issues are farther downstream and relate to the consequences of the knowledge generated by a large-scale project and the impact of its findings on individuals, groups, and society. Whether these longer range issues will actually arise in the future is difficult to predict, but it is, nonetheless, important to identify them and urge policymakers to be attentive to the project's potential longer term effects. In many areas, developing mechanisms for gathering broad public input will be critical, if not essential, to the project.

ISSUES RELATED TO RESEARCH POLICIES

- Does the United States have something unique to contribute or gain by sponsoring its own large population project?
- Will we learn important scientific, medical, and public health information that would otherwise be foregone if the United States did not sponsor a study? For example, does the United States have a unique genetic population or environment that is not found in other countries that are currently sponsoring studies?
- What is its value and cost?
- Given that the long-term cost required to mount a large population initiative will be significant, if not unprecedented, will it be possible to sustain public and political support for such an investment, especially since such support will need to be reaffirmed annually as part of the federal budget process?
- What would be the effect of funding the project on other areas of research or programs?
- Can existing studies achieve the same goals?
- Should there be collaboration with other countries conducting similar projects?
- Which agencies should be involved? Who should be the lead agency?
- What should be the role of the private sector?
- What intellectual property policies should govern the study—that is, who will own the rights to use or disseminate any discoveries resulting from the study?

Based on the input provided to NIH during its solicitation for input from the scientific community and on testimony provided to SACGHS during its public meetings, it is clear that there is a diversity of expert opinion within the scientific community about the wisdom of proceeding with a project of this magnitude at this time. For example, although a large cohort project may be needed to collect sufficient data to elucidate the contribution of genetic variation

and environmental factors to common diseases, some believe it may not necessarily lead to a better understanding of common diseases or population health benefits if it does not include a carefully designed, hypothesis-driven, disease-specific component. Others believe that such a study should *not* be hypothesis driven, but rather that it should be a viewed as a data resource for researchers to mine. Other concerns focus on the ripple effects such a costly study might have on other research areas or funding opportunities. Other questions arise about the need for collaboration with international efforts and the role of the private sector.

Need for and Value of Such a Project

- What evidence supports the inherent value of such a large population project?
- Will associations between genetic variation and common diseases be significant enough to bring about changes to clinical decisionmaking or lifestyle behaviors and to direct targeted public health education campaigns?
- Is the large cohort approach the only or most effective way to advance understanding of the interactions among genetics, environments, behaviors, their interactions, and common disease?
- Or, could existing cohort and case-control studies meet the same needs?

Some of the various arguments that have been put forward in response to these questions are outlined in the sections below.

Arguments in Favor of a Large Population Project. The basic premise underlying the enrollment of large numbers of participants as a resource for the study of human genetic variation and disease is the circumvention of the "signal-to-noise" problem present in smaller association studies. That is, small association studies may not have the statistical power to determine which of the many possible variables with perhaps small effect might be linked to common disease and truly contribute to health outcomes.

A new large population project is seen by its proponents as having the following advantages:

- control regarding the design of the data and specimen collection;
- the ability to collect and store biological data using the newest technologies;
- the use of a consistent, standardized protocol for collecting exposure, lifestyle, behavioral, and biological data;
- the ability to have a broad consent process that would cover all research needs; and
- the inclusion of diverse populations and ages not well represented in current studies.

Advocates of embarking on a large population project say that its major advantage is the increased statistical power it would have over traditional linkage studies and small association studies.[22] The need for greater statistical power is highlighted by the fact that there have been few successful attempts at identifying genes involved in common complex diseases with small

[22] Rosand, J., Altshuler, D. (2003). Human genome sequence variation and the search for genes influencing stroke. *Stroke.* 34:2512-2517.

cohorts in case-control association studies, and most attempts to replicate associations fail.[23] Without a large population cohort and rigorous statistical thresholds, some believe that association studies may be fraught with false-positive associations because of the lack of power needed to identify the multifactorial basis of common diseases and confirm initial findings through replication studies. Proponents of a large population project state that its scale would allow for the confirmation or refutation of existing hypotheses that would otherwise remain uncertain due to the constraints of the data currently available in smaller association studies.

An alternative to the large population project would be integrating existing prospective-cohort studies, such as the Framingham Heart Study, into the design of a new large population project. The argument against this option is that integrating the existing data could present several challenges.[24] Existing studies are limited in terms of the characteristics and phenotypes studied and the environmental exposures measured. Pooling could be difficult because of ascertainment biases, differing sampling strategies, missing data on critical variables, and survivor bias. In addition, existing studies have many more older participants and do not represent a full age range, and combining existing cohorts will not reflect the ethnic heterogeneity of the U.S. population.

The strengths of the design of a large population project include the replication of associations and the estimation of their magnitude, consistency, and temporality. Another benefit is that ultimately, the same population of participants, along with their genetic, medical, and exposure data, can be used to study the etiology of many common diseases. It may be difficult to obtain these benefits from existing cohort studies, particularly because most such studies focus on ascertaining only one or a few common diseases. Conducting multiple, individual case-control studies ultimately may require the same number of participants and resources. In a large population project, the links between various common diseases, such as hypertension and obesity, could be studied.

A collaboratively planned and implemented project within the United States may increase the participation of populations that are currently under-represented in research and the subsequent analysis of their genetic variations and disease risk. A U.S. project would include populations that are not represented in current international studies, such as African Americans, Latinos, Asians, and Native Americans. In addition, there will be a need to determine the environmental or exposure risks associated with disease in Americans that may require the collection of detailed behavioral, exposure, and sociocultural (e.g., poverty, education, diet) data on U.S. populations, as well as information about behavior and social conditions. Finally, although many existing biobanks intend to make their data available to researchers outside their countries, it is possible that access to international data and biological materials will be limited.

All of the potential benefits will need to be balanced against the expected cost of the project, its implications for other areas and types of research, and the potential burden on the subject

[23] Hirschhorn, J.N., Lohmueller, K., Byrne, E., Hirschhorn, K. (2002). A comprehensive review of genetic association studies. *Genetics in Medicine*. 4(2):45-61.

[24] National Institutes of Health (2005). *Summary of Public Responses to Request for Information: Design and Implementation of a Large-Scale Prospective Cohort Study of Genetic and Environmental Influences on Common Diseases*. See http://www.genome.gov/Pages/About/OD/ReportsPublications/ResponsestoRFINot-OD-04-041.pdf.

population, as well as the willingness of subjects to tolerate the questions and examinations required for participation.

Arguments That Favor Pooling of Existing Cohort Studies and Biobanks. Those who are more skeptical about the value of a large U.S. project argue that even by increasing the number of individuals in a cohort, the power of such a project still may not be sufficient to detect the gene-environment interactions correlated with disease due to the sheer number of strata present in multivariable conditions such as complex diseases. Thus, a large project may not facilitate the discovery of population health benefits.[25] Rather it might only suggest some associations that would have to be further pursued in smaller, more targeted studies. Some have suggested that pooling data and samples across multiple ongoing large population studies could be a reasonable alternative to mounting a new large-scale project and might be necessary anyway to avoid missing true associations.

Thus, as an alternative to establishing a new large population project, existing cohorts could be expanded to address the same experimental questions that a large project would address. The pooling of existing cohorts might save time and/or money because researchers could build upon existing DNA repositories, environmental exposure assessments, and infrastructure, rather than leaving these cohorts relatively underfunded for genetic analyses and their samples unanalyzed. The experience and expertise needed to execute an epidemiological project of this size already is available in groups that have undertaken existing studies. This experience can be drawn on to create relationships with the subject community and develop mechanisms for maintaining community trust in the study. The cost of a new project might jeopardize the ability of existing studies to maintain funding for their projects. In addition, existing studies have established long-term relationships with participants in which trust already has been formed. Investigators for these studies have detailed knowledge and invaluable experience with particular cohorts, which could be shared and built upon.

Another alternative to a single large project could be developed through a world-wide collaboration of existing large population biobanks. Because genetic markers do not change with time, the use of existing DNA repositories could reduce costs and allow genetic discovery projects to begin almost immediately. By using existing collections, resources could be dedicated to genotyping, dataset creation, and statistical analyses rather than cohort assembly. This type of collaboration would increase the statistical power to detect gene-environment interactions and could reduce errors in which false-positive associations are found. Such collaborations and their extremely large samples would be particularly valuable for allowing the subsetting of cases into independent groups in order to allow for the replication and confirmation of findings.
In sum, given current tight budgets and the failure to continue adequate support for existing cohorts, some members of the scientific community are uncertain about the wisdom of beginning a cohort project study of this size.

Arguments That Favor a Combination of Approaches. Some respondents to NIH's request for information on this topic argued against mutually exclusive options. Instead, they pointed out

[25] Khoury, M. (2004). The case for a global human genome epidemiology initiative. *Nature Genetics.* 36(10):1027-1028.

that although new cohorts are needed to supplement and extend existing cohorts, some existing cohorts could be supplemented to generate timely short- and medium-term data. In addition, existing cohorts would form valuable adjunct sources of data and undoubtedly provide more detailed data in certain areas than could the proposed project. They could be used to validate and extend findings from the large cohort and vice versa.

Some respondents noted that it may be wise to use existing cohorts for the purposes of the larger project in order to refine and pilot the processes of standardizing exposure, phenotype determination, and subject recruitment methods.

Arguments That Question the Value of a Large Population Project. It is possible that large population studies may not necessarily clarify the interaction between genes, environments, and phenotypes.[26] Pembrey and colleagues argue that, "Quite simply, the proper study of multifactorial traits or disorders demands the analysis of all the likely multiple factors in the same subjects over time. It is a question not so much of gene loci of small effect, but more of loci of *contingent* effect."[27] A study of social, behavioral, physiological, and physical exposure factors in the widest sense may need to be undertaken as part of a comprehensive general population (pre) birth prospective cohort study. However, by the time any prospective study is completed and the environmental influences are understood, the environments themselves will have "moved on"—that is, the environmental factors that were studied may not necessarily still have the same degree of social relevance or importance.

In addition, *a priori* ideas of the underlying mechanisms of disease are not necessarily hypothesized in large population studies, and, therefore, spurious results may be obtained by screening large numbers of potential etiologic factors for correlations with multiple diseases. This is in contrast to case-control studies in which a particular disease is identified and correlations with exposure or genetic variants are sought. Some in the scientific community worry that if the project does not involve *a priori* hypotheses about environment-gene-behavior, gene-disease-behavior, or environment-disease-behavior interactions, results may be suspect. It also is possible that too few clinically relevant events may accrue in a large population project, making the detection of genetic or environmental factors of small effect very difficult. Because common diseases are thought to be heterogeneous genetically and environmentally, the multiple influences underlying disease will require sample sizes large enough, and analytic techniques sensitive enough, to detect the multiple combinations of contributing genetic and environmental factors to the same phenotypic disease. Some scientists are not convinced that we have the necessary experience, infrastructure, or scientific culture in which to responsibly carry out a large and important project such as this. They argue that the genetic science of common complex diseases is simply not mature enough to launch such a costly initiative.

[26] Foster, M.W., Sharp, R.R. (2005). Will investments in large-scale prospective cohorts and biobanks limit our ability to discover weaker, less common genetic and environmental contributors to complex diseases? *Environmental Health Perspectives.* 113(2):119-122.

[27] Pembrey, M., ALSPAC Study Team (2004). The Avon longitudinal study of parents and children (ALSPAC): a resource for genetic epidemiology. *European Journal of Endocrinology.* 151:U125-U129.

Cost and Effects on Other Areas of Science

When the Human Genome Project was first conceived in the late 1980s, the scientific community was skeptical about its value and its possible negative impact on other areas of science, particularly hypothesis-driven, investigator-initiated science (typically funded through the R01 grant mechanism).[28] But as each phase of the project was completed and the promise realized, basic scientists became the greatest supporters of the project because it actually added value to and helped advance basic research.

Some parts of the scientific community are likely to have similar reservations and questions about another large-scale and costly scientific project. For example, what would happen to success rates if R01 funds were cut in order to fund this project? Will the emphasis on genetics and genomics detract attention and resources from other areas of research such as neuroscience or cell biology?

Any large population project will take years to become useful and will require enormous levels of funding to develop and maintain for years into the future. Depending on the funding mechanisms, this may require the widespread public support of government funding. The funding for such a project also may necessitate extensive private-public partnerships and collaboration, which may raise questions regarding commercialization.

Although there are no publicly available estimates of what such a project would cost, the figure could be as high as $3 billion, and perhaps higher.[29] Thus, if a tenth of the total ($300 million) were allocated each year for 10 years, care would have to be taken to ensure that no harm was done to other critical research and training programs. Some members of the scientific community have expressed concern about the impact that such a large allocation could have during flat funding periods and argue that a large project should be undertaken only if funded through sources that do not compromise investigator-initiated projects. They urge that there be careful consideration to balance the cost of a project versus other priorities either within HHS in general or within the biomedical research community. If tradeoffs must be made, how can they be accomplished with the least disruption to other promising areas of science?

Others have asked, because of the high costs of such a project, whether it is appropriate for the federal government to be the sole sponsor. Does the government have the necessary infrastructure to carry out this type of effort, or should it rely on the private investment of funds and technical resources to complete some of the work? If such a public-private partnership is developed, who would have rights to control access to the knowledge developed through use of the information generated by the project?

[28] Alberts, B. (1985). Limits to growth: in biology, small science is good science. *Cell.* 41:337-338.

[29] The $3 billion figure is a very rough, and probably conservative, estimate based on the projected long-range cost of the smaller National Children's Study, which was reported to be an estimated $3.2 billion. Testimony to House Appropriations Committee, Subcommittee on Labor, Health and Human Services, Education, and Related Agencies, Hearing on the National Institutes of Health on April 6, 2006.

Capacity to Conduct Interdisciplinary Science

Given the many scientific and academic disciplines that will be required to develop such a complex and broad-based project (as well as the current boundaries in academia) and the need for the involvement of teams of experts in human genetics, medicine, behavior, sociology, public health, epidemiology, and environmental health sciences, will there be insurmountable challenges to fostering such a multidisciplinary team approach?

Experience with other large-scale population projects has shown that it takes a tremendous amount of effort to agree on what should be measured, how to measure it, and how to interpret and translate the results. Some members of the scientific community are concerned that more work is needed to link the perspectives and methods of multiple disciplines and develop an interdisciplinary outlook before embarking on a large and expensive project.

Need for Partnerships

- In order to avoid duplication of efforts, and to ensure comparability of results, will a U.S. project need to collaborate with similar projects under way in other countries?
- If so, will one U.S. agency be designated to serve as the lead in interacting with the international efforts, and, if so, which agency?
- What is the appropriate role of the private sector in funding and implementing a large-scale project?

Representatives of international biobanks and large cohort projects reported to SACGHS that there is an increased need for harmonization in, for example, the taxonomy, techniques, and methods used to collect data and specimens. Others are calling for uniformity in the ethical principles and standards used in conducting large population-based projects. For example, the Secretary General of the United Nations stated, "Despite the existence of numerous declarations, guiding principles, and codes dealing with the issue of genetic data, the changing conditions of genetic research call for the establishment of an international instrument that would enable states to agree on ethical principles, which they would then have to transpose into their legislation." This need for harmonization is particularly important because human genetic research databases will be resources for many future uses.

With regard to the issues associated with the need for harmonization of the taxonomy, techniques, and methods used to collect data and specimens, SACGHS is aware of efforts currently under way to establish greater standardization in the technical aspects of data and specimen repositories. For example, Public Population Project in Genomics[30] is a non-profit organization that currently is building an international consortium to promote the type of discussions and collaborations needed to reach a consensus on optimal, standard procedures in the field of population genetics research.

[30] See www.p3gconsortium.org/.

Access to Data and Materials and Intellectual Property Concerns

Given that critical data and important discoveries are likely to emerge from such a project, some members of the scientific community have expressed concern over whether the data and materials that are derived from existing large-scale longitudinal studies will be widely available to the scientific community. Experience with existing similar population-based studies has shown a trend of limited access. Would this trend continue or would there be a policy of free and open access to this very expensive and valuable resource? In contrast to the limits placed on access to information generated from existing large population cohort studies, other more recent policies have focused on ensuring open and rapid access to data generated from, for example, the Human Genome Project and the HapMap Project. For example, the Bermuda Rules, drafted in 1996, state that "all human genomic DNA sequence information, generated by centers funded for large-scale human sequencing, should be freely available in the public domain in order to encourage research and development and to maximize its benefit to society."[31] This model of openness and access should be considered when setting policies for use of and access to data and materials gathered and/or analyzed as a result of a large population project. It will be critical for investigators to have access to data and specimens to cross validate markers and accelerate the clinical utility of the knowledge emerging from the project. In devising a policy for sharing data and specimens, it will be critical to institute procedures for protecting subject confidentiality. In addition, it is likely that some specimens and data will be particularly valuable because of their uniqueness or informational value, and precautions must be taken to ensure that specimens, data, and the rights of subjects are protected.

In addition, given the increasing rate at which genomic and proteomic discoveries are being patented, it will be important to clarify up front how ownership of intellectual property will be determined. The prospect of patent thickets or restrictive licensing of patents by institutions conducting the research could raise obstacles to the rapid development of public health measures associated with findings arising from the project.

Finally, the collaborative nature of the project will require new mechanisms of authorship recognition, given that scientific publications will potentially include the efforts of teams of scientists. The current promotion system in academia emphasizes independence, rather than teamwork, and it would need to be redesigned to recognize the contributions of individuals participating in a team effort.

Options for Addressing Research Policy Issues

In considering the issues involved in undertaking such a large-scale research project, the Secretary should ensure that there is widespread consultation about its merit and implementation with the U.S. scientific community, HHS agency leadership, the international community, and policymakers in Congress.

3. **The HHS Secretary, in consultation with the NIH Director, should ensure that there are opportunities available to the general scientific community to a) be informed**

[31] See www.ornl.gov/sci/techresources/Human_Genome/research/bermuda.shtml.

about the potential for such a project; b) present its views about the scientific validity and feasibility of such a project; c) present its views on the commitment of resources to such an effort, including whether there are benefits to leveraging existing efforts; and d) provide input on issues related to fair access by scientists to the resources and the sharing of data and samples.

4. **Given the transdisciplinary nature of the project and its potential scope, the Secretary may wish to establish a highly collaborative model of project leadership and management in multiple HHS agencies that includes biological, behavioral, social, public health, and population-scientific disciplines as well as basic biological scientists and epidemiologists.**

5. **The HHS Secretary should continue to promote and facilitate ongoing consultation with the international community and the private sector to explore opportunities for collaboration.**

6. **In embarking on such a large-scale project, the HHS Secretary, in consultation with the NIH Director, other HHS agencies, and appropriate congressional committees, should ensure that there is widespread support for sustaining a long-term and stable investment in a large population project.**

7. **To ensure that the public benefits from such discoveries, the Secretary should require that there be clear intellectual property policies in place for discoveries made using the data and samples collected through the project.**

ISSUES RELATED TO RESEARCH LOGISTICS

- How will the representativeness of the population be identified, defined, and achieved?
- Given that benefits to participants may only be indirect ones, will it be difficult to recruit a broad range of individuals?
- What are the ramifications of using racial or ethnic categories?
- Will the uninsured or underserved be part of the project, and, if so, how will they be recruited and maintained in the project?
- How will non-genetic study variables, such as environmental, social, physiological, and behavioral factors, be identified, defined, and studied?
- Will the lack of uniform methods for collecting, storing, and centralizing genetic, behavioral, social, physiological, and clinical health information make a project of this scale difficult to implement?
- Will new technologies be required to collect the necessary range of environmental data?
- When studies are performed using the data and materials collected in the project, to whom will findings of clinical significance be sent?

Beyond the specific design issues of a large-scale project are logistical considerations that could have social and ethical consequences. These considerations are worthy of public input and include developing enrollment and data collection procedures that accurately and fairly capture the ethnic, racial, and socioeconomic diversity of populations; coordinating across multiple

healthcare systems that lack universal, electronic medical record keeping systems; coordinating the multitude of healthcare institutions involved with the enrollment and data collection components of the project; creating new databases for data storage; defining and harmonizing the variables to be collected; and developing technologies to accurately and unobtrusively collect sociobehavioral and environmental information.

Enrollment Criteria and Recruitment of Racial/Ethnic Groups

In developing the criteria for the sample population to be recruited into a large project, several factors must be considered, not the least of which is the representativeness of that population. Enhancing representativeness will improve the likelihood that the results of the research can be broadly applied. However, defining representativeness is more complex than it might appear on the surface, and ensuring that once it is defined it can be achieved poses an additional challenge.

Race, ethnicity, and sex. Race/ethnicity has been shown to be associated with disease risk. However, race has been described in both biological and social terms. Moreover, in focusing on racial or ethnic differences, the differences among persons of the same socially defined racial or ethnic group may be overlooked, and the contribution of social and environmental factors may not be fully appreciated. The importance of race to health, particularly for complex diseases, is controversial.[32]

Thus, in deciding which groups to recruit into a large population project, it will be important to view race and ethnicity within a social context in which biological differences can sometimes, but not always, be found. In addition, Charles Rotimi warned that the inclusion of ethnic labels and a concentration on common genetic variants in the HapMap project, "ran the risk that this first approximation of human population structure might be subsequently used to reinforce existing racial or ethnic categories."[33]

Nonetheless, with careful consideration of all the factors that might be in play in disease initiation and progression, racial and ethnic categories can be useful for generating and exploring hypotheses about environmental and genetic risk factors, as well as interactions among risk factors, for important medical outcomes. And, despite the complexities and care that must be taken in attributing phenotypic differences to genetic differences among races, there is much to be gained by focusing on disorders that occur more frequently within a well-defined population. This clustering of disorders is not unusual among closely affiliated racial/ethnic groups, as it reflects the recent common ancestral origin, heritage, history, and environmental exposure of individuals within the group. It can be explained by the four forces of genetic drift, founder effects, selection, and the occurrence of new mutations. Ethnicity can be particularly important in determining some environmental influences, such as nutritional factors.

Thus, just how to categorize individuals for the purposes of recruiting subjects into a large project has both scientific and social implications. The ethical and social implications of studies

[32] Ioannidis, J.P., Ntzani, E.E., Trikalinos, T.A. (2004). Racial differences in genetic effects for complex diseases. *Nature Genetics*. 36(12):1312-1318. Epub 2004 Nov 14.

[33] Rotimi, C.N. (2004). Are medical and nonmedical uses of large-scale genomic markers conflating genetics and 'race'? *Nature Genetics Supp*. 36(11):S43-S47.

of human genetic variation include the possibility of confusing the socially informed definition of race or ethnicity with biologically derived definitions of populations. This could lead to inappropriate clinical interpretations as well as the over interpretation of meaning in ways that could stigmatize or harm entire groups of people.

Recruiting. There are several issues to consider in ensuring that the subject population is representative of the entire population, not just along gender, racial, or ethnic lines, but also along socioeconomic strata. Because the benefits of investing in such a project are likely to come in the future and not be uniformly shared across all participants, it might be more difficult to recruit certain groups of individuals who perceive the burdens and/or risks of participation as outweighing the potential benefits, given the other circumstances of their lives, such as poverty, discrimination, and lack of access to healthcare and other essential services. Furthermore, if entry into the project is through healthcare providers, what efforts will be made to ensure that those who lack access to healthcare are represented in the cohort? If enrollment is through healthcare providers, a major segment of the American population, the uninsured, will be under-represented. In addition, efforts will be needed to ensure that a representative sample is taken of individuals who rely on alternative medicine for their health and wellness care.

Measuring and Understanding Differences in Health and Risk Factors in the Population

One of the reasons we do not have a good sense of how genetic differences as opposed to environmental factors account for health differences is that we do not have adequate methodologies and technologies to measure of physical, physiological, and social environments. As such, it will be necessary to develop methods to collect data on diet and lifestyle, the initiation and progression of disease, and physiological and biobehavioral biomarkers. The techniques that might be needed to gather such data could be highly labor intensive, imposing potentially burdensome and intrusive requirements for subjects.

The Human Genome Project required the development of high-throughput data sequencing and computational tools to assemble, compare, and analyze digital data. A large cohort project demands the identification of a population that has sufficient breadth and depth to allow the analysis of myriad relevant questions, the identification of numerous biological variables to be measured—and their tabulation—and the creation of robust assessment and computational tools to define, measure, and assess the effects of environmental changes (including behavioral and social factors) over time. Compared to the Human Genome Project, these perceived requirements are far more complicated.

One particularly complex issue is the need to collect and analyze environmental exposure data. Not only would data have to be collected in the personal environments of subjects but also in the larger environments in which they live. Collecting accurate and reliable exposure data is extraordinarily difficult, requiring the development of new technologies.

As part of HHS's new GEI initiative, funds will be spent on developing new technologies to determine how the environment, diet, and physical activity contribute to illness. Investments will be made in emerging technologies, such as small, wearable sensors that can measure environmental agents that make contact with the body and individual measures of activity.

Devices also will be developed that measure changes in human biology, which can be observed in samples of blood or urine. Overall, it is hoped that these new tests will provide the precision needed to help determine how these factors influence the genetic risk of developing disease.

In addition to genes and the physical environment (i.e., quality of air, water, and housing conditions), health is also affected by behavioral, cultural, social and other factors. Moreover, these multiple factors may co-exist and interact to affect health and disease and can vary depending on stage of life, the extent and duration of exposure, and individual response. Gathering data and measuring the relative influence of these factors is important, complex, and labor intensive.

Coordination across Multiple Institutions and Healthcare Systems

The collection and dissemination of personal medical information in a large population project will be complicated in the United States, given that currently no universal electronic system for storing medical records exists. Furthermore, the enrollment of subjects will be influenced by the characteristics of a healthcare system that does not include uninsured, and often underinsured, Americans. In addition, another logistical challenge is posed by the fact that health insurance is primarily employer based, and many study subjects may change plans and providers frequently over the course of the project. A large population study in the United States could be difficult because it will put pressure on the American healthcare system, which is characterized by uncoordinated, decentralized private and public institutions.[34] Thus, given the current fragmented state of healthcare in this country, some members of the scientific community are asking whether a truly coherent cohort study can be designed, data collected and analyzed, and benefit returned to the participants and others at a reasonable cost.

Options for Addressing Research Logistics Issues

8. **To ensure diversity and appropriate representation in the population to be studied, the HHS Secretary should encourage project leadership and the scientific community to develop clear, consistent definitions and parameters for the stratification and classification of the projected sample population.**

9. **To ensure that subject selection is fair and just, the HHS Secretary should seek input from the public as well as researchers and clinicians on the best approaches to identifying subpopulations for recruitment, as well as issues to be considered in approaching, educating, and enrolling various subpopulations. Project organizers could be encouraged to consult with community-based organizations as one of many appropriate recruitment and enrollment strategies.**

10. **To refine methods for collecting and analyzing environmental (physical, behavioral, and social) factors influencing health, the HHS Secretary, in consultation with the NIH Director, should ensure that resources are devoted to developing new tools to**

[34] Altshuler, J.S., Altshuler, D. (2004). Organizational challenges in clinical genomic research. *Nature.* 429:478-481.

validate existing methods, as well as to improving assessments of the environment, as broadly defined.

11. **To develop uniform and secure approaches for collecting, storing, tracking, and centralizing clinical information to be gathered over the course of the project—including the use of electronic medical records—the HHS Secretary should encourage project leadership to consult with healthcare providers and organizations.**

ISSUES RELATED TO REGULATORY AND ETHICAL CONSIDERATIONS

- What are the regulatory requirements for the project, and how will they be met?
- Are there unique informed consent considerations?
- Will the project provide healthcare to its uninsured participants? If so, at what additional cost? And, will the project provide care that is only related to the goals of the project (e.g., initial intake examination)?
- If children or adolescents are to be enrolled, what additional protections must be considered?
- Who will have access to study data and biospecimens, under what circumstances, and how?
- Will the project require special arrangements or practices to enable participants to control how their samples and data are used?
- Will the project be able to accommodate participants' expectations regarding the confidentiality of their data?
- Will additional protections be necessary to reassure participants that their data will not be shared with health insurers, law enforcement agencies, or others?
- How and for how long will research data and samples be stored? How will they be secured? When and how will they be destroyed?
- Will study results emanating from the project be returned to participants, and what criteria will be used to determine when it is appropriate to return results?
- What federal, state, and local laws and regulations will need to be considered in deciding whether to return (or withhold) results emanating from the project to participants and/or their family members? Will investigators ever be faced with a duty to warn dilemma, and will policies need to be developed to anticipate such situations?
- How will results emanating from the project that could be relevant to non-participating family members be handled?

Many of the ethical and regulatory issues likely to be relevant to such a large-scale project are not necessarily unique or new. Rather, the size and magnitude of the project has the potential to either amplify or mask ethical concerns. It will be important for collection sites, data and specimen managers, and investigators to conduct activities consistently and uniformly and in accord with all ethical and regulatory requirements.

Institutional Review Board Review

Any large, epidemiological study of human genetic variation will require an ethics committee to review the ethical, legal, and social issues. A major challenge will be to coordinate multiple Institutional Review Boards (IRBs) for human subjects protections.

Several federal agencies play a role in defining and regulating the legal and ethical requirements for research involving human subjects, including banking human biological materials and clinical data. The HHS regulations at 45 CFR 46 provide the department's policy for the protection of human subjects. Also, within HHS, some research might be regulated by the Food and Drug Administration (FDA) if the protocols involve a clinical investigation regulated by FDA, or if it supports an application for research or marketing of an FDA-regulated product—for example, most drug, biologic, and device studies. Both sets of regulations stipulate that human subjects research must be reviewed by an independent body (an IRB) and that subjects must provide their informed consent to participate unless the requirements for informed consent are waived by an IRB according to the regulations. Typically, the IRB is situated locally to ensure that local community standards and norms are considered when research is reviewed.

A large population project could involve as many as 500,000 subjects at multiple (possibly hundreds) of sites. Several models exist for the use of a central IRB to provide consistent oversight across all sites, but some institutions remain resistant to non-local IRB oversight. Will all research sites in a large project have to agree to review by a central IRB, and how much modification can be made in review and approval procedures if local review is allowed?

In addition, depending on the inclusion criteria of the project, subjects who are considered vulnerable might be enrolled, such as individuals who are cognitively impaired or children. Special regulatory requirements must be met in the review of research involving these populations. What will the mechanisms be for ensuring that such requirements are met?

Informed Consent

At its simplest, informed consent must be effective and prospectively obtained, and the informed consent process involves three elements: 1) disclosing information to potential research subjects; 2) ascertaining that they understand what has been disclosed; and 3) ensuring their voluntariness in agreeing to participate in research.

The HHS and FDA regulations permit IRBs to approve research when informed consent is sought and documented from each prospective subject (45 CFR 46.111(a)(4)&(5); 21 CFR 56.111(a) (4)&(5)). Although the informed consent issues that would arise in the context of a large population project also appear in other types of research involving human subjects, several features of such a project deserve careful consideration with regard to obtaining informed consent from subjects.

The precise research purpose to which identifiable private information or specimens might serve may not always be known at the time of collection of clinical information, exposure data, or human biological materials, either because future studies have not yet been conceptualized or

because the data and samples might be made available to additional investigators at other institutions for different studies. Thus, individuals might not be told the specific use to which the specimens will be put. Some have argued that patients cannot adequately give consent unless they are provided with the specific details of each individual research study. However, recent studies suggest that many individuals find consent for future unspecified use of specimens acceptable.[35] Thus, in reviewing and approving a consent process to be used for a large population project, IRBs and investigators need to consider whether a one-time consent will be sufficient or whether periodic repetition or updating of the consent process should be required or considered.

A variety of informed consent protocols have been proposed to address situations in which future studies might require the use of data or samples for aims that were not originally identified at the time of donation and consent, including enlarged consent in which the established use of samples or data is modified in the future, consent with several options for research use, presumed consent, and blanket consent.[36] Developing appropriate and specific informed consent procedures will be challenging if all of the components of a project are not established at the outset.

Providing Care and the Therapeutic Misconception

There is considerable evidence that a major benefit of participating in a clinical trial or research study derives from the quality of general care provided by the research team, not just the experimental intervention. Ethicists have struggled to distinguish the investigator/subject relationship from the physician-patient relationship, because of concern about investigators' competing obligations to funders, their institutions, and science that may affect the care they can offer participants. Ethical complexities can arise when research subjects in protocols focused on diseases or conditions that affect them directly think of themselves as the recipients of healthcare services rather than as research subjects. The trust that potential subjects might place in the medical profession could affect their willingness to participate and their ability to provide informed consent free from undue influence.

Thus, critical considerations in designing the project and recruiting subjects include clarifying whether routine medical care will be provided as part of the project protocol and whether that creates an undue influence for participation (i.e., an individual with limited or no access to healthcare chooses to participate in order to receive care without fully appreciating the risks that might be involved). There will be many cases where the individual will fully appreciate the risks involved, but will need or want healthcare and will choose to participate on that basis regardless of the risks.

[35] Jack, A.L., Womack, C. (2003). Why surgical patients do not donate tissue for commercial research: review of records. *British Medical Journal*. 327(7409):262. See also Wendler, D. (2006). One-time general consent for research on biological samples. *British Medical Journal.* 332:544-547.

[36] Cambon-Thomsen, A. (2004). The social and ethical issues of post-genomic human biobanks. *Nature Reviews Genetics.* 5:866-873. In addition, previous advisory bodies have recommended approaches for crafting informed consent policies in human biological materials research. See, for example, National Bioethics Advisory Commission (1999). *Research Involving Human Biological Materials: Ethical Issues and Policy Guidance.* Rockville, MD: National Bioethics Advisory Commission. Available at www.georgetown.edu/research/nrcbl/nbac/hbm.pdf.

In addition, if during the course of research conducted using data and/or specimens collected through the project, the research results indicate the need for clinical intervention, what will be the responsibilities and obligations of the investigators to ensure that participants have access to the necessary care? It seems reasonable to conclude that the greater and clearer the health benefit to participants, the stronger the obligation.

Privacy and Confidentiality

Although there may be physical risks associated with medical procedures that might be used in a large population project (such as blood draws or biopsies taken for research purposes), risks to privacy and confidentiality also must be considered in research involving the long-term collection and storage of clinical data and human biological materials. This concern has increased as a result of advances in genetic and other molecular technologies. Research involving stored specimens can be conducted many years after specimen collection and has the potential to identify genetic or other molecular alterations that may have implications for the current or future health of subjects or their immediate family, such as the presence of disease or other unsuspected risks. In addition, the improper use or disclosure of such information could result in psychosocial harms (such as stigma) or the loss of employment or insurability.

Donors of biological samples and identifiable clinical information generally expect that their samples and data will be used in a way that advances knowledge or medical treatment but does not violate their privacy. In a study of hundreds of thousands of people, maintaining the confidentiality of participants' genetic information will be challenging. In fact, if a sufficient amount of an individual's DNA is sequenced, the genetic data become a unique identifier and, if additional information about the source of the sample is also available, the data could be used to identify the individual. For example, DNA sequences contained in a database could, in theory, be matched to identified individuals if additional biological samples were available from the identified individuals and if DNA sequence information from these identified samples is matched to DNA sequence data in the database. If further technological advances are made and genotyping becomes less expensive and routine, it may become easier to identify the source of DNA sequence data.[37] The protection of medical, exposure, and genetic information is critical for participants and groups who fear discrimination and stigmatization related to their genotypes. Varying levels of anonymity and coding schemes have been proposed to protect the identity of subjects, including the use of completely anonymous samples and data that cannot be traced to the identity of the subject. However, complete anonymity would hinder a study because it would prevent the collection of additional health status and exposure information in later phases. Also, in such a study complete anonymity is next to impossible, and study participants will need to be counseled accordingly.

In addition to the regulatory requirements for protecting the rights and welfare of human subjects, privacy regulations provide additional requirements regarding medical information. The Health Insurance Portability and Accountability Act of 1996 (HIPAA) Privacy Rule governs the protection of individually identifiable health information and was enacted to increase the privacy

[37] McGuire, A.L., Gibbs, R.A. (2006). No longer de-identified. *Science.* 312:370-371.

protection of health information with individual identifiers and to regulate known and unanticipated risks to privacy that may accompany the use and disclosure of such identified personal health information. It covers individually identifiable health information that is held or maintained by "covered entities" (health plans, healthcare clearinghouses, or healthcare providers who transmit health information for certain transactions as defined by HHS) or by business associates acting for a covered entity. The Privacy Rule does not apply to biological specimens per se, but it may apply to the identification of information associated with specimens. Thus, the Privacy Rule could have major implications for institutions participating in a large population project. In most cases, the Privacy Rule will require authorization from individuals (subjects) to use their protected health information in research, unless an exception applies. This authorization is distinct from informed consent, which is a separate process.

Control of Samples and Data

The issues of ownership of biological samples and research data and benefit sharing are relevant to the design of a large population project. Biological samples traditionally have been viewed as belonging to the researchers or institutions to which they were donated, and recent court rulings support the notion that individuals do not own their biological samples, regardless of whether a commercial benefit is expected from the research.[38] Cases in which subjects or patients have sued investigators or physicians for profiting from discoveries that were derived from the study of participants' samples have highlighted the need for appropriate informed consent procedures and ownership agreements. Some biobanks, such as the EGP, were intentionally designed to address benefit sharing by providing participants access to their genetic information and by establishing a private funding source. Participants have the right to access their personal genetic information for use in personalized medicine and the diagnosis of disease. Private research companies, rather than the government, fund the Estonian project, and they are given rights to subsequent pharmaceutical developments. Other projects, such as CARTaGENE in Canada and the Personalized Medicine Research Project at the Marshfield Medical Research Foundation, have been designed to exclude private ownership of samples and data or to return any profits from personalized medicine developments to the foundation.

Returning Research Results

There is ongoing debate about whether, and, if so, when, findings from research should be communicated to subjects—either upon completion of a study or at some later date. This issue is relevant to all research, not just to large population longitudinal projects. However, large population projects will be conducted over many years, if not decades, after the data and human biological materials were first collected, raising some additional questions about investigators' responsibilities to report potentially useful information to subjects.

Those who oppose revealing unanticipated and unconfirmed findings argue that the harms that could result from revealing preliminary data could be serious, including anxiety or unnecessary (and possibly harmful) medical interventions. Subjects could make burdensome, irreversible

[38] Hakimian, R., Korn, D. (2004). Ownership and use of tissue specimens for research. *Journal of the American Medical Association.* 292(20):2500-2505.

decisions based on information that ultimately may be proven false. On the other hand, those who believe that subjects have the right to have even interim research results cite the principle of autonomy, which dictates that subjects have a right to know what has been learned about them and the potential inaccuracy of that information.

The HIPAA Privacy Rule provides an individual the right of access to information about him- or herself, including personal research results obtained in the course of clinical care, with limited exceptions. The Privacy Rule not only gives patients a right to see their own records but also requires that patients be notified of their right to see such records. In addition, the Privacy Act of 1974 applies to certain personally identifiable information held by federal agencies in a "system of records" and thus applies to any research record held by HHS. Under this law, an agency must provide an individual access to his or her record. Moreover, the federal regulations for the protection of human subjects prohibit consent forms from including language "through which the subject or his representative is made to waive or appear to waive any of the subject's legal rights" (45 CFR 46.116).

These regulatory requirements could lead to an increase in the number of subjects who are aware of and who exercise their right to request and receive research results. Investigators will have to be prepared to include, and IRBs to review, plans for how to respond to subjects' requests for disclosure of research findings. Relevant questions include the following: How persuasive is the evidence of the validity of the results? What are the clinical implications of the research results? If the results could be used to treat or prevent serious disease, the impetus to return the results might be stronger.

Some research results might be considered clinically significant for some time. For example, research conducted on tissues today may be difficult to interpret for clinical purposes but could be meaningful in the future. Is the investigator responsible for reporting that information to identifiable subjects who participated in the research years ago? Clearly, in the clinical context it is the utility and validity of the information that should dictate a decision to contact patients with results. It is less clear whether an investigator, who has no therapeutic relationship with the subject, has the same obligation.

Another important requirement must be considered in the decision to report research results to subjects—that is, the Clinical Laboratory Improvement Amendments of 1988 (CLIA). CLIA regulations, which are enforced by CMS, do not permit the return of research results to patients or subjects if the tests were not conducted in a CLIA-certified laboratory. Thus, if a research laboratory is not CLIA certified, it should not be reporting results to subjects. In some circumstances, repeating the test in a CLIA-certified laboratory may be feasible and appropriate. In cases in which the CLIA regulations do apply to research laboratories, these laboratories may disclose test results or reports only to "authorized persons" as defined by state law. Most state laws do not include patients/research subjects as "authorized individuals" who may receive test results. However, some states, such as New Jersey and New York, do consider patients to be "authorized individuals," and other states, such as Connecticut and New Hampshire, stipulate that patients may receive results only with the permission of the ordering physician.

OPTIONS FOR ADDRESSING REGULATORY AND ETHICAL CONSIDERATIONS

12. To ensure that all research sites involved in the project are aware of and implement the regulations established to protect research subjects, medical privacy, and patient safety, the HHS Secretary, in consultation with the NIH Director, should convene a working group of representatives from the Office for Human Research Protections, FDA, the Office for Civil Rights, relevant HHS agencies, and the IRB and scientific communities to develop a set of recommended best practices and standard operating procedures for the project. Public input on the policies and procedures also can be sought.

13. To ensure that the appropriate protections of subjects' rights and welfare are in place and are being consistently implemented, project leadership should systematically and regularly seek the input of study subjects regarding their experiences, concerns, and recommendations for enhancing protections.

14. To promote the ethical use of clinical and epidemiological data and specimens collected through the project, project leadership could develop guidance on how such data and samples can be used and under what conditions. This guidance should be made available to project participants so that they are informed of the protections that are in place and that are to be expected.

ISSUES RELATED TO PUBLIC HEALTH IMPLICATIONS OF THE PROJECT

- Will the project's statistical genetic associations (or gene-environment associations) be robust enough to lead to new therapeutic or preventive strategies that are evidence based?
- Will such a project widen the gap between what can be diagnosed (or predicted) and what can be treated (or prevented)?
- Will data gathered at the broad population level be applicable to all communities and groups?
- How will study results emanating from the project, which may magnify the complexity of population risk assessment, be implemented by regulatory health and safety agencies?
- Do regulatory agencies, local public health departments, and healthcare providers have sufficient resources to translate the knowledge that such a project will generate into clinical practice?

From statistical standpoint, studies have shown that there is sufficient power to detect the presence of causative polymorphisms of small effect with as few as 500 individuals are sampled. Greater power is achieved by increasing the sample size. In a large population project, there will be a desire to use all of the power of the sample size to highlight definitive findings and statements that, although reflective of that population, are actually representative of the local heterogeneity of the genetic/environmental factors. Moreover, it is incredibly difficult in many cases to move from a statistical genetic association to an understanding of the mechanism of action that would suggest new therapies or preventive measures or that would withstand evidence-based regulatory decisionmaking. This leads to questions about whether large population cohort studies can actually provide results that are sufficiently definitive to lead to

clinical applications and whether the data gathered can be reliably extrapolated across the entire population.

Without the ability to identify gene function, there also is the risk that genes or SNPs will be associated with disease, but we will not know with what certainty. Or, the association will be clear, but no treatment would be available. This gap between identifying risk and providing treatment is troublesome, particularly because of its uncertain duration. Thus, it will be critical to employ a variety of mechanisms to disseminate information resulting from the research, such as conferences, publications, and public fora.

OPTION FOR ADDRESSING THE PUBLIC HEALTH IMPLICATIONS OF THE PROJECT

15. To advance the application of research findings resulting from the project to improve health, the HHS Secretary and project leadership should systematically and regularly disseminate study findings as they emerge from the project, with clear descriptions of the possible clinical implications of the results and the limitations of the data, their generalizability, and their clinical and public health implications. This information should be tailored to meet the information needs of the public, healthcare providers, and the public health community.

ISSUES RELATED TO SOCIAL IMPLICATIONS OF THE PROJECT

- Given the range of genetic and environmental factors involved in common disease, could such a project create new health disparities or change the way we currently think about them?
- Could the findings resulting from the project exacerbate existing vulnerabilities such as age, race, and disability?
- If research findings using project data and specimens result in the identification of new vulnerable populations, will there be sufficient social and public health resources available to respond?
- If the project generates clinically useful knowledge, will it largely benefit only those with access to the healthcare system?
- Can the project results be realized in a decentralized and fragmented healthcare system?
- Could the findings from such a project exacerbate racial discrimination and group stigmatization?
- What are the views of minority communities about the project's implications?
- Will the project pose risks of genetic discrimination, given the lack of comprehensive legal protections at the federal level?
- Could research findings lead to simplistic and reductionist explanations of the role of genetics in disease and result in the misinterpretation of genetics in public policy, in the courts, and the in the provision of health and life insurance?

Elucidating and/or Exacerbating Health Disparities

Eliminating health disparities is an important goal of public health efforts. It is unclear the extent to which genetic differences account for health disparities because most current genetic studies do not have adequate measures of the physical and social environment. Health disparities research provides an important opportunity to integrate biological knowledge with social/behavioral knowledge in order to better understand the determinants of disease, which will help us to reduce the risk of disease and to provide better treatment when it arises. In the context of a large population project, it is necessary to study the multiple risk factors simultaneously within subgroups (e.g., race, ethnicity, behaviors, geography, genetic backgrounds, exposures, and social environments) to understand how environmental and genetic risk factors interact and lead to health differences. Thus, large population cohort studies could help clarify or change the way we think about health disparities. For example, research may determine that a particular group of individuals (e.g., a specific racial or ethnic group) has an increased risk of developing disease. However, if this particular group also is socially or economically vulnerable, the findings could exacerbate disparities as well as discriminatory practices.

Rotimi and others caution that because of the continuity of variation across all human populations, race and other culturally derived notions of groups should not be used as a proxy for genotype when scientists study variation and physicians diagnose and treat genetic diseases.[39] Culturally defined groups, such as African Americans, could experience stigmatization if race is used as a delimiting factor in characterizing genetic variation. For example, if a particular genetic variant that predisposes one to a particular disease is prevalent in one subpopulation, the medical community and public may naively assume that all members of the race to which the given subpopulation belong are predisposed to having the disease.

Consultation with leaders in the groups likely to be affected may be one of many helpful ways of informing decisionmakers. Consultation with leaders of potentially/historically vulnerable groups—for example, racial/ethnic groups, women, gay and lesbians, those with lower education attainment—will be particularly critical.

The Risks of Genetic Determinism

The belief that genes alone determine everything about an individual is called "genetic determinism." However, although genes play an essential role in the formation of physical and behavioral characteristics, each individual is the result of a complex interaction between his or her genes and the environment within which he or she develops, beginning at the time of fertilization and continuing throughout life. As social and biological beings, we are creatures of our biological, physical, social, political, historical, and psychological environments. The great lesson of modern molecular genetics is the profound complexity of gene-gene interactions, gene-environment interactions, and gene-environment-behavior interactions in the determination of whether a specific trait or characteristic is expressed in an individual.

[39] Royal, C.D.M., Dunston, G.M. (2004). Changing the paradigm from 'race' to human genome variation. *Nature Genetics Supp.* 36(11):S5-S7.

Although the concept of complete genetic determinism is can be overly simplistic, genes do play a role in determining biological characteristics, including a predisposition to certain diseases. However, recent scientific findings have revealed that a "one-gene, one-disease" approach is far too one dimensional. Knowing the complete DNA sequence of a gene, even one on the relatively small list of genes currently associated with a specific disease, does not allow a scientist to predict whether a given person will get the disease. And, even when a specific genetic change is identified that "causes" the disease in some people, others may be found who have the same change but do not develop the disease. This is because other factors—genetic, epigenetic, environmental, or behavioral—are altered that mask or compensate for the disease gene. Thus, even with the most sophisticated understanding of genes, one cannot always determine with certainty what will happen to a given person with a single change in a single gene. This means that there is a danger that misrepresentation of risks could occur, not just among the lay public, but also among professionals and policymakers. In addition, given that most genetics research is still focused on identifying single causative factors and has not matured to complex models of genetic causation, scientists themselves sometimes promote a naive biological, deterministic interpretation of complex disorders. This is likely to lead to further misinterpretation and misuse of these genetic explanations in public policy, the courts, health and life insurance policies, and medical practices.

Recent efforts to address concerns about the potential for genetic discrimination in the workplace and in the health and life insurance industries have met with mixed success. Although a 2000 Executive Order prohibits genetic discrimination in federal employment, no such protection exists for non-federal workers. Legislation extending this protection to all workers passed the U.S. Senate most recently in 2005, but a House version has failed to reach a vote. Until such protection is universally available to all Americans, the scientific community must be extra vigilant in ensuring that research results are appropriately interpreted and communicated to the public.

Developing Reasonable Social and Policy Responses to Research Findings

Genetic findings in complex disorders, especially gene-environment interactions, are not likely to provide enough of an impetus to allow regulatory bodies to create policies to protect people. Although there has been some progress lately in the field of gene-environment interactions, such as in toxicogenomic and pharmacogenomic research, the results themselves have exposed the immense complexity that is involved in integrating this type of knowledge into existing policy standards and methods.

Traditionally, public health policy has focused on population-level solutions—a one-size-fits-all model—such as can be seen in some anti-smoking campaigns. Nobody would disagree with the promotion of smoking cessation as a population public health effort. In contrast, genetic information is based on the individual and the family and even on ethnic groups and will require intense research efforts on the implications of the use of specialized policies and regulations for the protection of vulnerable populations. For example, what if it is found that some people are sensitive to their environments and others are not? Will it become the responsibility of the "sensitive" individual to take him- or herself out of harm's way or have behavioral interventions

imposed on him or her so that the rest of society can ignore the vulnerability or impugn ultimate responsibility for it?

The current risk-assessment paradigm used by the regulatory agencies (e.g., EPA, FDA) is population based. How will these agencies set standards and guidelines for businesses and products based on complex susceptible genetic subgroups? And, will they have the resources to make the changes in their regulatory paradigms?

OPTION FOR ADDRESSING THE SOCIAL IMPLICATIONS OF THE PROJECT

16. To periodically assess persistent and emerging social implications of the project and research results, the HHS Secretary, in consultation with project leadership, should establish an independent standing committee for the duration of the project. The committee could consist of individuals with expertise in the relevant sciences, medicine, law, ethics, and patient and community advocacy. The committee should routinely seek public input on the implications of the research resulting from the project and report its findings.

V. OPTIONS FOR ENGAGING THE PUBLIC

With the growing enterprise of clinical research has come the need to inform members of the public about the underlying science, engage them in discussions of priorities for federal research spending, and seek support for important areas of research. However, new issues with strong scientific content sometimes seem particularly ill-suited to one-time techniques for soliciting opinion (e.g., a typical opinion poll).[40] Because most members of the public will be unfamiliar with the concepts of a large population project, concerted efforts must be made to educate, inform, and solicit feedback and input. Over the last 10 to 15 years, increasing efforts to consult lay people about scientific issues have produced a range of new methods for doing so.

The public can be consulted at many levels. At one end of the spectrum, policymakers can choose simply to inform or educate the public. A more consultative approach to public engagement assumes that the public brings to the issue and the topic experiences and perspectives and values that will help inform overall policy. In addition, there are many "publics," including the general public, disease advocacy groups, scientific and professional organizations and their members, healthcare providers, and healthcare organizations.

Mechanisms for public engagement can include polling, surveys, moderated focus groups, workshops, or scenario development. Issue identification and agenda setting can rest with the organizers or, by contrast, to obtain public input, the organizers can ask the community to help identify issues, frame and prioritize the discussion, and devise outreach strategies. A more deliberative approach to engaging the public involves providing in-depth background information about the topic to better facilitate public formulation of what the issues are. In the case of a large population project and its research uses, engagement may be aimed toward the

[40] *Information and Attitudes: Consulting the Public About Biomedical Science* (2005). A report published by the Wellcome Trust.

communities from which participants will be recruited, or it may require a more national or regional conversation.

To have a credible deliberative process, participation must be broad and representative. The information that is presented should be balanced, accurate, and fair, and the process and settings need to be such that a safe and ample opportunity exists for everyone to hear and to be heard. Equally as important, the policymakers and decisionmakers need to be a part of the process from its onset to demonstrate to participants that their time and effort is worthwhile.

Project leaders will need to recognize that public engagement is enhanced when various groups recognize the relevance of the project to the public. Communities must understand the purpose of the project, how it is designed, and who will benefit from it. The project process must be explicit in addressing the issues of individual and group stigma and the representativeness of racial, ethnic, and other groups, and it must be sincere and provide real and meaningful ways of involvement in developing plans and methods. However, once community expectations are raised, not fulfilling these expectations can lead to mistrust and opposition. Community-based organizations can serve as one of many valuable intermediaries that can help initiate and maintain consistency. Moreover, a model of community-based participatory research can offer some relevant lessons for a large population project. The knowledge gained is bidirectional, going from the community to the researchers and from the researchers to the community. This can build trust and ensure participation. If the project leadership consults with the community and is fully participatory, it can become a vehicle for community education as it moves forward.

According to the public engagement experts SACGHS has consulted, the proposed project poses the potential of generating mistrust, especially among members of under-represented racial and ethnic communities. However, some recent research has demonstrated small differences in the willingness of minorities to participate in health research compared to non-minority populations.[41] Ensuring that these groups support the project may depend on how successfully the concept of co-ownership is embraced across the communities that perceive that they are most at risk from the project. If a sense of co-ownership is achieved, however, powerful advocates will support the building of the infrastructure for the project that will be necessary. Decisionmaking and planning must engage the community from the outset. The process should be explicit in addressing the issues of race and racism, and the individual representatives of racial and ethnic groups must be meaningfully involved in developing project plans and methods. In addition, national organizations representing racial/ethnic populations should be engaged.

In planning and implementing the four large population studies that are under way in Iceland, Estonia, the United Kingdom, and Quebec, public engagement was an important part of the process, whether it occurred in a formal and structured manner (through a legislative or regulatory process) or through a planned program of outreach to the public that included seeking public opinion and inviting comments (see Appendix B).

[41] Wendler, D., Kington, R., Madans, J., et al. (2006). Are racial and ethnic minorities less willing to participate in health research? *PLoS Med.* 3(2):e19.

Serious efforts at public engagement are likely to employ a mixed strategy—the various methods of addressing the public are not mutually exclusive. The goals should be to raise awareness, educate, obtain feedback, and establish a relationship with the public or segments of the public. The public engagement process used to incorporate public input into recommendations for pandemic influenza vaccination priorities may serve as a model. The report can be found at www.keystone.org/spp/health-pandemic.html.

What has to be decided is the point at which public consultation will be sought (see Figure A).

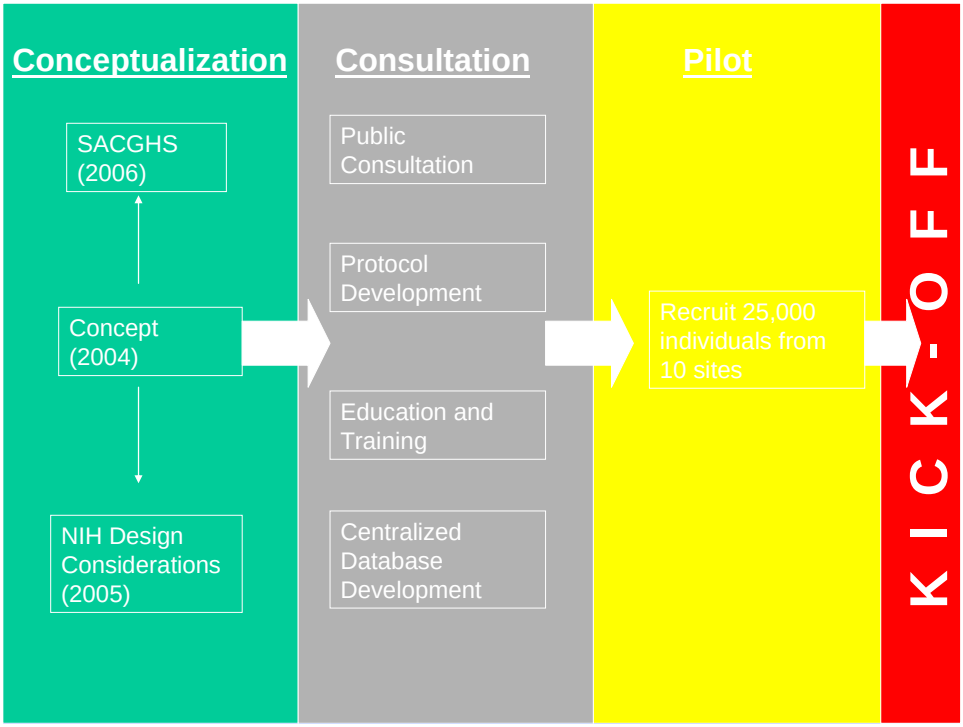

Figure A: Steps in Public Consultation

That is, should the public be consulted to inform a "go/no-go" decision on the project, once protocol development begins, after implementation, or as an ongoing process? In addition, about what should the public be consulted? (see Figure B).

Should they be asked to provide feedback on the informed consent process, sample collection procedures and processing, and data access? Should health professionals be consulted about how best to interpret and report research results? Should healthcare organizations be consulted about how to maintain privacy and confidentiality while sharing data and specimens?

Whatever strategies are selected for ensuring public engagement, project leaders will need to be prepared to consider and use the information and feedback provided and, if necessary, revise project goals or design and initiate additional rounds of consultations.

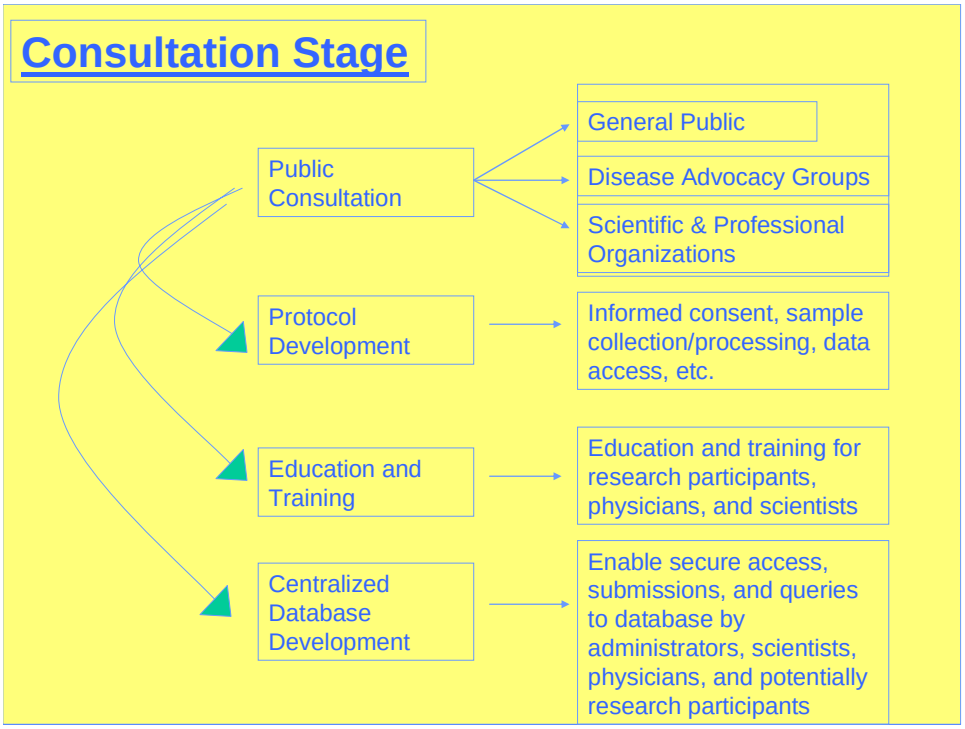

Figure B: Stages of Public Consultation

NHGRI Public Consultation Initiative

On February 14, 2005, NHGRI announced the availability of funds for conducting a pilot public consultation study to obtain wide societal input to inform the design and implementation of one or more possible large U.S. population-based studies, including a longitudinal cohort study, of the role of genes and the environment in health and disease. Funds totaling $2.1 million are to be awarded in FY 2006 and FY 2007 to one specialized center. The project will be funded using a cooperative agreement funding mechanism that will enable NHGRI to partner with the awardee in carrying out the project.

The aim of the project is to solicit opinions on the design and implementation of the study from members of the public representative of the demographic makeup of the country. The issues to be addressed may include, but are not necessarily limited to, the following:

- the acceptability of goals of the initiative for the United States as a whole;
- concerns regarding the uses of data, for individuals, communities, and the public at large;
- expectations about privacy protection;
- the acceptability of open-ended consent;
- the acceptability of a central IRB;
- optimal approaches to recruitment, particularly regarding identifying and contacting family members;
- the need for tailoring to individuals or communities with special needs;

- expectations about the return of information to individuals, communities, and the public at large;
- the need for ongoing dialogue with participants regarding study goals and processes;
- the advisability of including or excluding children; and
- intellectual property concerns.

The pilot project will obtain public input through several methods, including surveys, focus groups, and public meetings. The methods are to be proposed by the applicants, but NHGRI will collaborate in designing the survey instruments, focus group guides, and final protocols for the public meetings.

The findings of each of the elements of the pilot public consultation study are to be analyzed as they proceed; an overall analysis of the findings also is to be conducted. A preliminary analysis of the data is to be completed by September 2008 and incorporated into the design of the longitudinal cohort study, its full-scale public consultation component, and other population-based studies, should they be determined to be feasible and should they be funded within the next few years. If a large-scale study proceeds after the completion of the pilot project, additional consultation would take place, with the specific communities to be recruited.

OPTIONS FOR PUBLIC ENGAGEMENT

The Committee encourages efforts to be made at all levels to develop a broader understanding of the issues involved so they can be identified early in the process and addressed fairly and responsibly, both before and throughout the duration of the proposed project. In previous sections of this report, SACGHS has suggested several options for engaging the public in discussions and decisions about undertaking a large population project, including consulting with the scientific and international communities, communities that might be involved in the research, healthcare providers and their institutions, and those who volunteer to participate in the project as research subjects.

1. **The public's willingness to participate in a large population project should be assessed before embarking on such an expensive endeavor. Willingness could be assessed through opinion polls, requests for comments posted on agency websites, or through other measures. Such an assessment should be made in advance of a funding decision.**

2. **If a decision is made to proceed with the project, it will be important to ensure that public engagement occurs throughout all aspects and stages of the research process, from conceptualization through design, planning, implementation, conduct, and data analysis and reporting. Public engagement also will be important in applying the knowledge gained by the research and in addressing its implications. The Secretary should ensure that sufficient project resources are dedicated to public consultation activities both before and throughout the duration of the project.**

VI. CONCLUSION

SACGHS's goal is to help illuminate a pathway for the HHS Secretary's assessment of the merit, utility, and feasibility of a large population project. Although in this report SACGHS has identified considerable challenges, the Committee is enthusiastic about the concept of mounting a large population project for the study of genes, environments, their interactions, and common diseases in the United States because of its potential to generate significant health benefits.

APPENDIX A
INTERNATIONAL BIOBANKING EFFORTS

U.K. Biobank

The U.K. Biobank's main aim is to elucidate the effects of genetic and environmental factors on the risk of common multifactorial diseases of adult life. The bank aims to enroll 500,000 male and female middle-aged participants and is the largest and most ambitious biobank established to date. The U.K. Biobank directors believe that a large sample is needed to provide the statistical power necessary to conclusively detect meaningful correlations of phenotype, genotype, and environmental exposure and to identify the multiple factors of often modest effect that contribute to disease. Its goal is to establish a prospectively gathered collection of samples, in conjunction with comprehensive measures of exposure and phenotype, so that a wide range of gene/exposure/phenotype relationships can be studied. The U.K. Biobank directors intend for the biobank to serve as a resource for the biomedical research community for decades to come. Each participant will contribute a blood sample, complete a questionnaire on lifestyle, provide a medical history, and receive an examination by a nurse. Participants also will be followed regularly through their physicians for the reporting of morbidity and disease diagnosis and will be resurveyed periodically to update their exposure data. Blood samples will be stored for future retrieval for nested case-control studies. This information will be made available to researchers investigating the complex interactions among genes, environment, and lifestyle that are believed to cause many complex disorders, such as cancer, heart disease, diabetes, and Alzheimer's disease. Researchers will apply through a peer-review mechanism for access to data and for approval of their research proposal to investigate a specific disease. To ensure that the identity of gene donors is protected, scientists or medical doctors who request data will be given de-identified, coded data.

Biobank Japan

Biobank Japan is a fully funded national project designed to collect blood samples from 300,000 Japanese residents, with the goal of developing personalized medicine for a set of 40 diseases, including cancer, diabetes, rheumatoid arthritis, and other common disorders.[42] The samples will be genotyped by the Human Genome Center of the Institute of Medical Science of the University of Tokyo with a production-scale SNP BeadLab (Illumina, Inc.), a state-of-the-art tool for the analysis of genetic variation and function. A new facility to house the specimens will be built, and specimens will be kept separate from medical and genetic data. To ensure that the identity of gene donors is protected, samples will be coded. An explicit informed consent process will be utilized.

[42] Triendl, R. (2003). Japan launches controversial Biobank project. *Nature Medicine*. 9(8):982.

Estonian Genome Project

The Estonian Genome Project (EGP) is a biobank of information on diseases, lifestyle, demographics, genealogy, and DNA in a database that is accessible by researchers. These data might be used to inform the clinical treatment of sample donors and for public health research and statistical applications. The data are free of charge to academic researchers. Foreign researchers can use the data in collaboration with Estonian scientists. The project plans to enroll 100,000 of Estonia's 1.4 million people. In addition to advancing the development of diagnostics and therapies, EGP is designed to be of direct benefit to participants and allows study participants to access their own genetic information. For example, should scientists discover that a genetic variant is an indicator of a particular disease or an indicator of an adverse reaction to medication, a donor can request information on his or her genotype at the particular variant. To ensure that the identity of gene donors is protected, scientists or medical doctors requesting data are given de-identified, coded data.

Icelandic Genetics and deCODE

In 2000, the Icelandic Parliament granted deCODE, a pharmaceutical company, exclusive rights to the country's medical records. These records exist in the Icelandic Health Sector Database. At the same time, Iceland's government authorized deCODE to begin the construction of a biobank of the Icelandic population. Iceland's population is a unique resource because it is composed of a homogenous population for which an extensive genealogical database, including information on previous generations dating back hundreds of years, is available. Iceland's population of 275,000 has been geographically isolated over time. Its genetic homogeneity, limited population, and extensive genealogical information are frequently cited by deCODE as the major reasons for its success in discovering new genes and genetic material related to several common diseases.

Because this extensive genealogy is available for the entire population, individuals from extended families who have the same disease can be grouped and studied using linkage analysis to identify the biomarkers and segments of particular chromosomes. These segments of chromosomes are likely to contain genes related to the disease that can serve as targets for pharmaceutical development. The approach taken by deCODE is augmented by the country's extensive and well-developed system of medical records that have been well maintained since 1915. The identities of the participants in the deCODE studies are kept secret through use of an encrypted identification code. To date, deCODE has identified genes involved in several common complex diseases, including myocardial infarction, stroke, osteoporosis, and asthma.

APPENDIX B
PUBLIC CONSULTATION IN INTERNATIONAL PROJECTS

Within the four large population genetic studies that are under way in Iceland, Estonia, the U.K., and Quebec, two major approaches are being taken to consult with the public and gather opinion. Godard et al. classify the type of consultation being conducted in Estonia and Iceland as focused more on quantitative rather than qualitative data.[43] In embarking on these consultations, deCODE and the Estonian government did not specifically "reach out" to the public, but the launching of these biobanks and the rules governing them were established through the legislative process, which includes discussion.[44] The U.K. and Quebec, on the other hand, are engaging the public in a "participation or partnership approach," with a focus on both quantitative and qualitative measures of public opinion.[45]

deCODE and the Icelandic Healthcare Database

deCODE looked at community consent as a necessary prerequisite to performing a large cohort DNA study. Iceland's Parliament passed a law allowing the development of the Icelandic Healthcare Database (IHD), which can be viewed as proof of community consent. The debate that occurred took place through hundreds of newspaper articles and television programs and several town hall meetings across the country, and it informed the passage of the law, affecting the database license that was granted by parliament. According to deCODE, "debate is one of the most important mechanisms by which complex ideas are processed by democratic societies."[46] Following the debate and the media coverage, a survey indicated that 75 percent of the Icelandic population supported passage of the bill to allow the IHD. A survey taken in 2000 after the law was passed indicated that support by the public had grown to 90 percent (although this second survey might have used misleading wording/content and may not be accurate).[47,48] Organized and vocal dissent to the legislation still occurs, and more than 10 percent of the Icelandic population has chosen to opt out of the study, indicating that support for the project may be weaker than surveys have suggested and that the public consultation did not successfully address the concerns of a significant portion of the population. One opposing group, Mannvernd, has brought attention to some public concerns over human rights and private control of medical and genetic information that are not addressed in the legislation allowing the IHD.

[43] Godard, B., Marshall, J., Laberge, C., Knoppers, B.M. (2004). Strategies for consulting with the community: the cases of four large-scale genetic databases. *Science and Engineering Ethics*.10(3):457-477.
[44] Working Party on Biotechnology (2005). *Tokyo Workshop Report: Human Genetic Research Databases—Issues of Privacy and Security*. Organisation for Economic Co-operation and Development, DSTI/STP/BIO (2005)14.

[45] Godard, B., et al. (2004). Op. cit.
[46] Gulcher, J., Stefánsson, K. (2000). The Icelandic Healthcare Database and informed consent. *The New England Journal of Medicine*. 342:1827-1830.
[47] Ibid.
[48] Godard, et al.(2004). Op. cit.

Estonian Genome Project

Estonia used a similar approach to that of Iceland for consultation. The test of public opinion has been limited to Gallup poll results.[49] The Estonian Genome Project (EGP) informed the public about the pertinent scientific information regarding the project, and public support/approval was based on this educational process. The project website provides information, definitions, and news stories related to the project.[50] There appears to be less opposition to the EGP than to the deCODE project in Iceland, but this is likely because of some aspects of the projects themselves (i.e., Estonia's project carefully addresses concerns in the areas of consent, confidentiality, trust, and discrimination), rather than because of public education and consultation efforts.[51] In addition, the EGP received more media attention than did the deCODE project during the stages before the legislation was enacted. Since the inception of the program, Estonia has commissioned a marketing and consulting company to conduct polls to assess knowledge about and opinion on the project.[52]

U.K. Biobank

The funders of the U.K. Biobank acknowledge the importance of consulting with the public as instrumental to the project's success and as valuable to the shaping of policies and practices.[53] The Wellcome Trust and the British Office of Science and Technology believe that the "engagement model," of dialogue between scientists and the general public is a better form of public consultation and communication than the "deficit model," which merely provides information about science to the public.[54] The idea for the U.K. Biobank first appeared in 1999, and by 2000 the first public consultation was undertaken. The Biobank consultations were preceded by reports from the Wellcome Trust on public views of science and consultations and the role of scientists in public consultation and debate.[55] The first Biobank-specific consultation focused on the public perceptions of human biological sample collection.[56] In establishing principles to govern this collection, in the context of the large population cohort, the Wellcome Trust/Medical Research Council framework for collection was discussed with spokespeople for certain public groups and with scientists (those with an interest in medical research). Sixteen focus groups, composed of a diverse range of members of the general public, were formed to address policy concerns surrounding biological sample collection. Factors such as ethnic group, age, socioeconomic group, and geographic location were taken into account to ensure adequate representation of the entire population.[57] The topics the focus groups looked into included awareness of, understanding of, and attitudes about topics such as medical and genetic research and human biological samples. Questions about deciding to donate, anonymity and

[49] Ibid.
[50] Ibid.
[51] Ibid.
[52] Working Party on Biotechnology. Op. cit.
[53] Science and the Public: A Review of Science Communication and Public Attitudes to Science in Britain. A Joint Report by the Office of Science and Technology and the Wellcome Trust. October 2000.
[54] Ibid.
[55] Ibid.
[56] U.K. Biobank, available at www.ukbiobank.ac.uk/ethics/consultations.php.
[57] The Wellcome Trust and the Medical Research Council (2000). *Public Perceptions of the Collection of Human Biological Samples*. London. Available at www.phgu.org.uk/ecard?reference_ID=3870.

confidentiality, consent, and ownership of these samples were addressed.[58] In addition to these focus groups, in-depth interviews with specific stakeholder groups (including medical professionals, individuals or family members of people with a disease or disability, and community and religious leaders) were conducted.[59] This initial consultation was followed by several subsequent consultations; these included consultation with primary care health professionals on the recruitment of patients; consultation with social groups that were under-represented at the initial consultation; a workshop with medical professionals, social scientists, patient advocates, lawyers, ethicists, and civil society groups to discuss ethics; consultation with industry representatives; a public panel (of previously consulted individuals without a stake in the project) on governance and framework; and a workshop with stakeholders on governance and framework.[60] The second round of consultations, with healthcare professionals, looked at developing the protocol for the project. The consultations that followed focused on oversight and ethical concerns in the following areas: feedback, access to the database, and withdrawal from the study.[61] In total, the U.K. Biobank sponsored 12 different consultations with various public groups.

Reports from each of these consultations, detailing the objectives, methods, findings, and more, can be found on the U.K. Biobank website at www.ukbiobank.ac.uk/ethics/consultations.php.

CARTaGENE

In establishing its biobank, CARTaGENE in Quebec hopes to engage the public in a partnership decisionmaking process.[62] The first stage of their communication with the public—focus groups that included members of Quebec's population (randomly selected from the phonebook[63]) representing its diverse linguistic, cultural, and regional groups—were held to look at the social and ethical implications of the CARTaGENE project and the social perceptions of the project.[64] The primary goal of this stage of the consultation was to identify the concerns of the public regarding the establishment of this biobank.[65] In November 2001, four preliminary sets of these of focus groups were held to gauge the popular opinion of this project.[66] In the fall of 2003, 19 of these focus groups (of 7 to 8 people each) were held to obtain a larger scale view of the social and ethical concerns of the public. This initial set of focus groups was followed by a large-scale survey developed to assess how true the results of the focus groups were to the general public. This survey was conducted in all regions of Quebec, with more than 1,300 people agreeing to

[58] Ibid.
[59] Ibid.
[60] U.K. Biobank website at www.ukbiobank.ac.uk/ethics/consultations.php.
[61] Working Party on Biotechnology. Op. cit.
[62] Godard, B. (2004). *CartaGene (Abstract). Genome Canada GE^{3}LS Project Presentations*. Accessed Online August 2, 2005, at www.genomecanada.ca/ge3ls2005/proceedings/08_04.asp.
[63] Godard, B., Marshall, J., Laberge, C., Knoppers, B.M. (2004). Strategies for consulting with the community: the cases of four large-scale genetic databases. *Science and Engineering Ethics*.10(3):457-477.
[64] Godard, B. (2003). *Consulting Communities: A Matter of Trust and Communication* (Presentation). Accessed Online August 2, 2005, at www.humgen.umontreal.ca/genconsult/docs/9.pdf.
[65] Godard, B., Marshall, J., Laberge, C., Knoppers, B.M. (2004). Strategies for consulting with the community: the cases of four large-scale genetic databases. *Science and Engineering Ethics*.10(3):457-477.
[66] Godard, B. (2003). *Consulting Communities: A Matter of Trust and Communication* (Presentation). Accessed Online August 2, 2005, at www.humgen.umontreal.ca/genconsult/docs/9.pdf.

participate.[67] The second stage of communicating with the public consisted of developing a plan for how to communicate with the public before and during the project. To do this, workshops with ethics, law, and policy experts were held to look at what sorts of communication were needed with the public before embarking on the project. During the summer of 2001, information was shared with the public through the CARTaGENE website, newsletters, and ongoing press releases and interaction with the media. In June 2003, a second workshop of professionals was held, and from six months before beginning recruitment through the project's initiation, a telephone hotline was set up to respond to the questions and concerns of the public, and information about the project was dispersed through fliers, posters, and the website.[68] The final stage of CARTaGENE's consultation involved the establishment of a "deliberative electronic forum," in which the public can discuss concerns and share opinions with researchers, allowing for a dialogue between the two groups that could continue throughout the study.[69]

[67] Godard, B., Marshall, J., Laberge, C., Knoppers, B.M. (2004). Strategies for consulting with the community: the cases of four large-scale genetic databases. *Science and Engineering Ethics.*10(3):457-477.
[68] Godard, B. (2003). *Consulting Communities: A Matter of Trust and Communication* (Presentation).
[69] Godard, B., Marshall, J., Laberge, C., Knoppers, B.M. (2004). Strategies for consulting with the community: the cases of four large-scale genetic databases. *Science and Engineering Ethics.*10(3):457-477.